Taylor Polynomial Estimator for Estimating Frequency Moments

Sumit Ganguly
Indian Institute of Technology, Kanpur
sganguly@cse.iitk.ac.in

Abstract

We present a randomized algorithm for estimating the pth moment F_p of the frequency vector of a data stream in the general update (turnstile) model to within a multiplicative factor of $1 \pm \epsilon$, for $p > 2$, with high constant confidence. For $0 < \epsilon \leq 1$, the algorithm uses space $O(n^{1-2/p}\epsilon^{-2} + n^{1-2/p}\epsilon^{-4/p}\log(n))$ words. This improves over the current bound of $O(n^{1-2/p}\epsilon^{-2-4/p}\log(n))$ words by Andoni et. al. in [2]. Our space upper bound matches the lower bound of Li and Woodruff [23] for $\epsilon = (\log(n))^{-\Omega(1)}$ and the lower bound of Andoni et. al. [3] for $\epsilon = \Omega(1)$.

Contents

1 Introduction

The data stream model is relevant for online applications over massive data, where an algorithm may use only sub-linear memory and a single pass over the data to summarize a large data-set that appears as a sequence of incremental updates. Queries may be answered using only the data summary. A data stream is viewed as a sequence of m records of the form (i, v), where, $i \in [n] = \{1, 2, \ldots, n\}$ and $v \in \{-M, -M+1, \ldots, M-1, M\}$. The record (i, v) changes the ith coordinate f_i of the n-dimensional *frequency vector* f to $f_i + v$. The pth moment of the frequency vector f is defined as $F_p = \sum_{i \in [n]} |f_i|^p$, for $p \geq 0$. The (randomized) F_p estimation problem is: Given p and $\epsilon \in (0, 1]$, design an algorithm that makes one pass over the input stream and returns $\hat{F}_p$ such that $\mathsf{Pr}\big[|\hat{F}_p - F_p| \leq \epsilon F_p\big] \geq 0.6$ (where, the constant 0.6 can be replaced by any other constant $> 1/2$.) In this paper, we consider estimating F_p for the regime $p > 2$, called the *high moments* problem. The problem was posed and studied in the seminal work of Alon, Matias and Szegedy in [1].

Space lower bounds. Since a deterministic estimation algorithm for F_p requires $\Omega(n)$ bits [1], research has focussed on randomized algorithms [5, 11, 31, 21, 32, 17, 23, 3]. Andoni et. al. in [3] present a bound of $\Omega(n^{1-2/p} \log(n))$ *words* assuming that the algorithm is a *linear sketch*. Li and Woodruff in [23] show a lower bound of $\Omega(n^{1-2/p}\epsilon^{-2} \log(n))$ bits in the turnstile streaming model. For *linear* sketch algorithms, the lower bound is the sum of the above two lower bounds, namely, $\Omega(n^{1-2/p}(\epsilon^{-2} + \log(n)))$ words.

Space upper bounds. The table in Figure 1 chronologically lists algorithms and their properties for estimating F_p for $p > 2$ of data streams in the *turnstile mode*. Algorithms for *insertion-only* streams are not directly comparable to algorithms for update streams—however, we note that the best algorithm for insertion-only streams is by Braverman et. al. in [7] that uses $O(n^{1-2/p})$ bits, for $p \geq 3$ and $\epsilon = \Omega(1)$.

Contribution. We show that for each fixed $p > 2$ and $0 < \epsilon \leq 1$, there is an algorithm for estimating F_p in the general update streaming model that uses space $O(n^{1-2/p}(\epsilon^{-2} + \epsilon^{-4/p} \log(n)))$ words, with word size $O(\log(nmM))$ bits. It is the most space economical algorithm as a function of n and $1/\epsilon$. The space bound of our algorithm matches the lower bound of $\Omega(n^{1-2/p}\epsilon^{-2})$ of Li and Woodruff in [23] for $\epsilon \leq (\log n)^{-p/(2(p-2))}$ and the lower bound $\Omega(n^{1-2/p} \log(n))$ words of Andoni et.al. in [3] for linear sketches and $\epsilon = \Omega(1)$.

Algorithm	Space in $O(\cdot)$ words	Update time $O(\cdot)$
IW[20]	$n^{1-2/p} \left(\epsilon^{-1} \log(n)\right)^{O(1)}$	$(\log^{O(1)} n)(\log(mM))$
Hss[6]	$n^{1-2/p}\epsilon^{-2-4/p} \log(n) \log^2(nmM)$	$\log(n) \log(nmM)$
MW [24]	$n^{1-2/p}(\epsilon^{-1} \log(n))^{O(1)}$	$n^{1-2/p}(\epsilon^{-1} \log n)^{O(1)}$
AKO[2]	$n^{1-2/p}\epsilon^{-2-4/p} \log(n)$	$\log n$
BO-I [8]	$n^{1-2/p}\epsilon^{-2-4/p} \log(n) \log^{(c)}(n)$	$\log n$
this paper	$n^{1-2/p}\epsilon^{-2} + n^{1-2/p}\epsilon^{-4/p} \log(n)$	$\log^2(n)$

Figure 1: Space requirement of published algorithms for estimating F_p, $p > 2$. Word-size is $O(\log(nmM))$ bits for algorithms for update streams. $\log^{(c)}(n)$ denotes c times iterated logarithm for $c = O(1)$.

Techniques and Overview. We design the Geometric-Hss algorithm for estimating F_p that builds

upon the Hss technique presented in [6, 15]. It uses a layered data structure with $L+1 = O(\log n)$ levels numbered from 0 to L and uses an ℓ_2-heavy-hitter structure based on CountSketch [12] at each level to identify and estimate $|f_i|^p$ for each heavy-hitter. The heavy-hitters structure at each level has the same number of $s = O(\log n)$ hash tables with each hash table having the number of buckets (height of table). The main new ideas are as follows. The height of any CountSketch table at level l is α^l times the height of any of the tables of the level 0 structure, where, $0 < \alpha < 1$ is a constant. The geometric decrease ensures that the total space required is a constant times the space used by the lowest level and avoids increasing space by a factor of $O(\log n)$ as in the Hss algorithm.

In all previous works, an estimate for $|f_i|^p$ for a *sampled item* i was obtained by retrieving an estimate $\hat{f}_i$ of f_i from the heavy-hitter structure of an appropriately chosen level, and then computing $|\hat{f}_i|^p$. In order for $|\hat{f}_i|^p$ to lie within $(1 \pm \epsilon)|f_i|^p$, $|\hat{f}_i - f_i|$ had to be constrained to be at most $O(\epsilon|f_i|/p)$. By the lower bound results of [26], the estimation error for CountSketch is in general optimal and cannot be improved. We circumvent this problem by designing a more accurate estimator $\bar{\vartheta}(\lambda, k)$ for $|f_i|^p$ directly. If λ is an estimate for $|f_i|$ that is accurate to within a constant relative error, that is, $\lambda \in (1 \pm O(1/p))|f_i|$ and there are independent, identically distributed and unbiased estimates $X_1, X_2, \ldots, X_{\Theta(k)}$ of $|f_i|$ with standard deviation $\sigma[X_j] \leq O(|f_i|/p)$, then, it is shown that (i) $\mathbb{E}\left[\bar{\vartheta}(\lambda, k)\right] \in (1 \pm O(1/p)^k)|f_i|^p$, and (ii) $\mathsf{Var}\left[\bar{\vartheta}(\lambda, k)\right] \leq O(|f_i|^{2p-2}\sigma^2[X_j])$.

The estimator $\bar{\vartheta}$ is designed using a *Taylor polynomial estimator*. Given an estimate $\lambda = |\hat{f}_i|$ for $|f_i|$ such that $\lambda \in (1 \pm O(1/p))|f_i|$, the $k+1$ term *Taylor polynomial estimator* denotes $\vartheta(\lambda, k) = \sum_{j=0}^{k} \binom{p}{j} \lambda^{p-j}(X_1 - \lambda)(X_2 - \lambda)\ldots(X_j - \lambda)$, where, $X_1, \ldots, X_k$ are independent and identically distributed estimators of $|f_i|$. Note that replacing the X_j's by $|f_i|$ gives the expression $\sum_{j=0}^{k-1} \binom{p}{j} \lambda^{p-j}(|f_i| - \lambda)^j$, which is the degree-$k$ term Taylor polynomial expansion of $|f_i|^p$ around λ (i.e., $(\lambda + (|f_i| - \lambda))^p$. A new estimator $\bar{\vartheta}(\lambda, k, r)$ is defined as the average of r *dependent* Taylor polynomial estimators ϑ's, where, each of these r ϑ-estimators is obtained from a certain k-subset of random variables $X_1, \ldots, X_s$, with $s = O(k)$, and each k-subset is drawn from an appropriate code and has a controlled overlap with another k-subset from the code. Note that now, only a constant factor (i.e., within a factor of $1 \pm O(1/p)$) accuracy for the estimate λ of $|f_i|$ is needed, rather than an $O(\epsilon)$-accuracy needed earlier.

Finally, we note that Hss algorithm [15] used full independence of hash functions and then invoked Indyk's method [19] of using Nisan's pseudo-random generator to fool space-bounded computations [25]. In our algorithm, we show that it suffices to use only limited $d = O(\log n)$-wise independence of hash families, by changing the way the hash functions are composed.

Notation

Let $\mathbb{R}$ denote the field of real numbers, $\mathbb{N}$ denote the set of natural numbers, that is, $\mathbb{N} = \{0, 1, 2, \ldots, \}$, $\mathbb{Z}$ denote the ring of integers, and $\mathbb{Z}^+$ and $\mathbb{Z}^-$ denote the set of positive integers and the set of negative integers respectively.

For $a \in \mathbb{R}$ and $s \in \mathbb{N}$, define

$$a^{\underline{s}} = \begin{cases} a \cdot (a-1) \cdot \cdots \cdot (a-s+1) & \text{if } s \in \mathbb{Z}^+ \\ 1 & \text{if } s = 0 \ . \end{cases}$$

It follows that, (i) for $s_1, s_2 \in \mathbb{N}$, $a^{\underline{s_1+s_2}} = a^{\underline{s_1}}\,(a-s_1)^{\underline{s_2}}$, and (ii) for $a < 0$, $a^{\underline{s}} = (-1)^s(-a+s-1)^{\underline{s}}$. The notation $a^{\underline{s}}$ is taken from [27].

For $p \in \mathbb{R}$ and $k \in \mathbb{N}$, denote

$$\binom{p}{k} = \begin{cases} \frac{p^{\underline{k}}}{k!} & \text{if } p \in \mathbb{R} \text{ and } k \in \mathbb{N} \\ 0 & \text{if } p \in \mathbb{R} \text{ and } k \in \mathbb{Z}^- \end{cases} .$$

We use the well-known following identities for binomial coefficients, namely, the *absorption identity*: $\binom{p}{k} = \frac{p}{k}\binom{p-1}{k-1}$, for integer $k \neq 0$, and, the *upper negation identity*: $\binom{p}{k} = (-1)^k\binom{k-p-1}{k}$, for integer k.

Review: Residual second moment and CountSketch algorithm

Let $f \in \mathbb{Z}^n$ and let rank : $[n] \to [n]$ be any permutation that orders the indices of f in non-decreasing order by their absolute frequencies, that is, $|f_{\text{rank}(1)}| \geq |f_{\text{rank}(2)}| \geq \ldots |f_{\text{rank}(n)}|$. The k-residual second moment of f is denoted by $F_2^{\text{res}}(k)$ and is defined as $F_2^{\text{res}}(k) = \sum_{i \in [n], \text{rank}(i) > k} f_i^2$.

We will use the CountSketch algorithm by Charikar, Chen and Farach-Colton [12], which is a classic algorithm for identifying ℓ_2-based heavy-hitters and for estimating item frequencies in data streams. The CountSketch(C, s) structure consists of s hash tables denoted $T_1, \ldots, T_s$, each having C buckets. Each bucket stores an $\log(nmM)$ bit integer. The jth hash table uses the hash function $h_j : [n] \to [C]$, for $j = 1, 2, \ldots, s$. The hash functions are chosen independently and randomly from a pair-wise independent hash family mapping $[n] \to [C]$. A pair-wise independent Rademacher family $\{\xi_j(i)\}_{i \in [n]}$ is associated with each table index $j \in [s]$, that is $\xi_j(i) \in_R \{-1, 1\}$. The Rademacher families for different j's are independent. Corresponding to a stream update of the form (i, v), all tables are updated as follows.

for $j = 1$ **to** s **do**
  $T_j[h_j(i)] = T_j[h_j(i)] + v \cdot \xi_j(i)$
endfor

Given an index $i \in [n]$, the estimate $\hat{f}_i$ returned for f_i is the median of the estimates obtained from each table, namely,

$$\hat{f}_i = \text{median}_{j=1}^{s} T_j[h_j(i)] \cdot \xi_j(i) .$$

It is shown in [12] using an elegant argument that

$$\left| \hat{f}_i - f_i \right| \leq \left(\frac{8F_2^{\text{res}}(C/8)}{C} \right)^{1/2} . \tag{1}$$

2 Taylor polynomial estimator

Let X be a random variable with $\mathbb{E}[X] = \mu$ and $\mathsf{Var}[X] = \sigma^2$. Singh in [29] considered the following problem: Given a function $\psi : \mathbb{R} \to \mathbb{R}$, design an unbiased estimator θ for $\psi(\mathbb{E}[X])$ (i.e., $\mathbb{E}[\theta] = \psi(\mathbb{E}[X])$. His solution for an analytic function ψ was the following. Let $\psi(t) = \sum_{k \geq 0} \gamma_k(0) t^k$. Let ν be a distribution over $\mathbb{N}$ with probability mass function $p_\nu(n)$, for $n = 0, 1, 2, \ldots,$. Choose $n \sim \nu$ and define the estimator

$$\theta = (p_\nu(n))^{-1} \gamma_n(0) \cdot X_1 \cdot X_2 \ldots \cdot X_n$$

where the X_i's are *independent copies of* X. The estimator satisfies

$$\mathbb{E}[\theta] = \sum_{n\geq 0}(p_\nu(n))^{-1}\cdot p_\nu(n)\cdot\gamma_n(0)\mathbb{E}[X_1]\,\mathbb{E}[X_2]\ldots\mathbb{E}[X_n] = \sum_{n\geq 0}\gamma_n(0)\mu^n = \psi(\mu)\ .$$

However, the variance can be large; for the geometric distribution ν with $p_\nu(n) = q(1-q)^n$, for $n \geq 0$ and $0 < q \leq 1$, it is shown in [10] that $\mathbb{E}\left[\theta^2\right] = (1/q)\sum_{n\geq 0}\gamma_n^2(0)((\mu^2+\sigma^2)/(1-q))^n$.

2.1 Taylor Polynomial Estimator

The Taylor polynomial estimator (abbreviated as TP estimator) is derived from the Taylor's series of $\psi(\mu) = \psi(\lambda + (\mu - \lambda))$ by expanding it around λ, an estimate of μ, and then truncating it after the first $k+1$ terms. Let $X_1, \ldots, X_k$ be independent variables with the same expectation $\mathbb{E}[X_j] = \mu = \mathbb{E}[X]$ and whose variance is each bounded above by σ^2. Define

$$\vartheta(\psi, \lambda, k, \{X_l\}_{l=1}^k) = \textstyle\sum_{j=0}^k \gamma_j(\lambda)(X_1-\lambda)(X_2-\lambda)\ldots(X_j-\lambda)\ .$$

where, $\gamma_j(t)$ is the function $\psi^{(j)}(t)/j!$, for $j = 0, 1, \ldots$. Its expectation and variance properties are given below. Let $\eta^2 = \mathbb{E}\left[(X_j-\lambda)^2\right] = \sigma^2 + (\mu-\lambda)^2$, for $j = 1, \ldots, k$.

Lemma 1. *Let $\{X_l\}_{l=1}^k$ be independent random variables with expectation μ and standard deviation at most σ. Let $\eta = (\sigma^2 + (\mu-\lambda)^2)^{1/2}$ and let ψ be analytic in the region $[\lambda, \mu]$. Then the following hold.*

1. *For some $\lambda' \in (\mu, \lambda)$, $\left|\mathbb{E}\left[\vartheta(\psi, \lambda, k, \{X_l\}_{l=1}^k)\right] - \psi(\mu)\right| \leq |\gamma_{k+1}(\lambda')|\cdot|\mu-\lambda|^{k+1}$.*
2. *$\mathsf{Var}\left[\vartheta(\psi, \lambda, k, \{X_l\}_{l=1}^k)\right] \leq \left(\sum_{j=1}^k |\gamma_j(\lambda)|\eta^j\right)^2$.*

Corollaries 2 and 3 apply the Taylor polynomial estimator to $\psi(t) = t^p$.

Corollary 2. *Assume the premises of Lemma 1. Further, let $\psi(t) = t^p$, $p \geq 2$, $\mu > 0$, $|\lambda - \mu| \leq \alpha\mu$, for some $0 \leq \alpha < 1/2$ and $k+1 > p$. Then,*

$$\left|\mathbb{E}\left[\vartheta(x^p, \lambda, k, \{X_l\}_{l=1}^k\right] - \mu^p\right| \leq \left(\frac{\alpha}{1-\alpha}\right)^{(k+1)}\cdot\mu^p\cdot\left(\frac{p}{k+1}\right)^{\lfloor p\rfloor+1}\ .$$

In particular, for p integral, $\mathbb{E}\left[\vartheta(x^p, \lambda, k, \{X_l\}_{l=1}^k\right] = \mu^p$.

Corollary 3. *Assume the premises of Lemma 1 and Corollary 2. Then*

$$\mathsf{Var}\left[\vartheta(x^p, \lambda, k, \{X_l\}_{l=1}^k\right] \leq (1.08)p^2\mu^{2p-2}\eta^2\ .$$

2.2 Averaged Taylor polynomial estimator

We use a version of the Gilbert-Varshamov theorem from [4].

Theorem 4 (Gilbert-Varshamov)**.** *For positive integers $q \geq 2$ and $k > 1$, and real value $0 < \epsilon < 1 - 1/q$, there exists a set $\mathcal{C} \subset \{0,1\}^{qk}$ of binary vectors with exactly k ones such that $\mathcal{C}$ has minimum Hamming distance $2\epsilon k$ and $\log|\mathcal{C}| > (1 - H_q(\epsilon))k\log q$, where, H_q is the q-ary entropy function $H_q(x) = -x\log_q\frac{x}{q-1} - (1-x)\log_q(1-x)$.*

Corollary 5. *For $k \geq 1$, there exists a code $Y \subset \{0,1\}^{8k}$ such that $|Y| \geq 2^{0.08k}$, each $y \in Y$ has exactly k 1's, and the minimum Hamming distance among distinct codewords in Y is $3k/2$.*

Let Y be a code as given by Corollary 5. Each $y \in Y$ is a boolean vector $y = (y(1), y(2), \ldots, y(s))$ of dimension $s = 8k$ with exactly k 1's. It can be equivalently viewed as a k-dimensional ordered sequence $y \equiv (y_1, y_2, \ldots, y_k)$ where $1 \leq y_1 < y_2 < \ldots < y_k \leq s$, and y_j is the index of the jth occurrence of 1 in y. Let $\pi : [k] \to [k]$ be a permutation and $y = (y_1, \ldots, y_k)$ be an ordered sequence of size k. Then, $\pi(y)$ denotes the sequence of indices $(y_{\pi(1)}, \ldots, y_{\pi(k)})$.

Let $X_1, X_2, \ldots, X_s$ be independent random variables with expectation μ and standard deviation at most σ. We first define the Taylor polynomial estimator, denoted TP estimator, for $\psi(\mu)$, given (i) an estimate λ for μ, (ii) a codeword $y \in Y$, and (iii) a permutation $\pi : [k] \to [k]$. The TP estimator corresponding to $y \in Y$ and permutation π is defined as

$$\vartheta(\psi, \lambda, k, s, y, \pi, \{X_t\}_{t=1}^{s}) = \sum_{v=0}^{k} \gamma_v(\lambda) \prod_{l=1}^{v} \left(X_{y_{\pi(l)}} - \lambda \right) \ .$$

Let $\{\pi_y\}_{y \in Y}$ denote a set of $|Y|$ randomly and independently chosen permutations that map $[k] \to [k]$ that is placed in (arbitrary) 1-1 correspondence with Y. The averaged Taylor polynomial estimator AVGTP averages the $|Y|$ TP estimators corresponding to each codeword in Y, ordered by the permutations $\{\pi_y\}_{y \in Y}$ respectively, as follows.

$$\bar{\vartheta}(\psi, \lambda, k, s, Y, \{\pi_y\}_{y \in Y}, \{X_l\}_{l=1}^{s}) = \frac{1}{|Y|} \sum_{y \in Y} \vartheta(\psi, \lambda, k, s, y, \pi_y, \{X_l\}_{l=1}^{s}) \tag{2}$$

The Taylor polynomial estimator in *RHS* of Eqn. (2) corresponding to each $y \in Y$ is referred to simply as ϑ_y, when the other parameters are clearly understood from context. Note that for any $y \in Y$ and permutation π_y, $\mathbb{E}[\vartheta_y]$ is the same. Therefore, due to averaging, the AVGTP estimator has the same expectation as the expectation of each of the ϑ_y's.

Lemma 6. *Let $p \geq 2, q = 8$, $k \geq \max(1000, 40(\lfloor p \rfloor + 2))$ and $s = qk$. Let $Y \subseteq \{0,1\}^s$ such that, (a) $|Y| \geq 2^{0.08k}$, (b) each $y \in Y$ has exactly k ones, and (c) the minimum Hamming distance among distinct codewords in Y is $3k/2$. Let $\{X_1, \ldots, X_s\}$ be a family of independent random variables, each having expectation $\mu > 0$ and variance bounded above by σ^2. Let λ be an estimate for μ satisfying $|\lambda - \mu| \leq \min(\mu, \lambda)/(25p)$ and let $\sigma < \min(\mu, \lambda)/(25p)$. Let $\eta = ((\lambda - \mu)^2 + \sigma^2)^{1/2} > 0$. Let $\bar{\vartheta}$ denote $\bar{\vartheta}(t^p, \lambda, k, s, Y, \{\pi_y\}_{y \in Y}, \{X_l\}_{l=1}^{s})$. Then*

$$\mathsf{Var}\left[\bar{\vartheta}\right] \leq \left(\frac{(0.288)p^2}{k} \right) \mu^{2p-2} \eta^2 \ .$$

3 Algorithm

The Geometric-Hss algorithm uses a level-wise structure corresponding to levels $l = 0, 1, \ldots, L$, where, the values of L and the other parameters are given in Figure 2.

Level-wise structures

Corresponding to each level $l = 0, 1, \ldots, L-1$, a pair of structures $(\mathsf{HH}_l, \mathsf{TPEst}_l)$ are kept, where, HH_l is a $\mathsf{CountSketch}(16C_l, s)$ structure with $s = O(\log n)$ hash tables each consisting of $16C_l$

Description of Parameter	*Parameter and its value*
Number of levels	$L = \lceil \log_{2\alpha} \frac{n}{C} \rceil$
Reduction factor	$\alpha = 1 - (1 - 2/p)\nu,\ \nu = 0.01$
Basic space parameters	$B = \left(\frac{425(2\alpha)^{p/2} n^{1-2/p} \epsilon^{-2}}{\min(\epsilon^{4/p-2}, \log(n))} \right)$ $C = (27p)^2 B$
Level-wise space parameters	$B_l = 4\alpha^l B, \quad l = 0, 1, \ldots, L-1$ $C_l = 4\alpha^l C, \quad l = 0, 1, \ldots, L-1$ $C_L = 16(4\alpha^L C),$
Degree of independence of $g_1, \ldots, g_L$	$d = 50\lceil \log n \rceil$
Taylor Polynomial Estimator Parameters	$k = 1000\lceil \log n \rceil$, $r = 16k, s = 8k$
Degree of independence of table hash functions	$t = 11$

Figure 2: Parameters used by the Geometric-Hss algorithm.

buckets. The TPEst_l structure is used by the Taylor polynomial estimator at level l and is a standard $\mathsf{CountSketch}(16C_l, 2s)$ structure with the following minor changes.

(a) The hash functions h_{lr}'s used for the hash tables T_{lr}'s are 6-wise independent.

(b) The Rademacher family $\{\xi_{lr}(i)\}_{i\in[n]}$ is 4-wise independent for each table index $r \in [2s]$, and is independent across the r's, $r \in [2s]$.

The hash tables $\{T_{lr}\}_{r\in[2s]}$ have $16C_l$ buckets each and use the hash function h_{lr}, for $r \in [2s]$. Corresponding to the final level L, only an HH_L structure is kept which is a $\mathsf{CountSketch}(C_L^*, s)$ structure, where $C_L^* = 16C_L$. The structure at level L uses $O(1)$ times larger space for HH_L to facilitate the discovery of all items and their frequencies mapping to this level (with very high probability).

Hierarchical Sub-sampling

The original stream $\mathcal{S}$ is sub-sampled hierarchically to produce random sub-streams for each of the levels $\mathcal{S}_0 = \mathcal{S} \supset \mathcal{S}_1 \supset \mathcal{S}_2 \supset \cdots \mathcal{S}_L$, where, $\mathcal{S}_l$ is the sub-stream that maps to level l. The stream $\mathcal{S}_0$ is the entire input stream. $\mathcal{S}_1$ is obtained by sampling each item i appearing in $\mathcal{S}_0$ with probability 1/2; if i is sampled, then all its records (i, v) are included in $\mathcal{S}_1$, otherwise none of its records are included. In general, $\mathcal{S}_{l+1}$ is obtained by sampling items from $\mathcal{S}_l$ with probability 1/2, so that $\mathsf{Pr}[i \in \mathcal{S}_{l+1} \mid i \in \mathcal{S}_l] = 1/2$. This is done by a sequence of independently chosen random hash functions $g_1, g_2, \ldots, g_L$ each mapping $[n] \to \{0, 1\}$. Then,

$$i \in \mathcal{S}_l \ \text{ iff } g_1(i) = 1, g_2(i) = 1, \ldots, g_l(i) = 1, \quad l = 1, 2, \ldots, L \ .$$

If $i \in \mathcal{S}_l$, then for each stream update of the form (i, v), the update is propagated to the structures HH_l and TPEST_l.

Group thresholds and Sampling into groups

Let $\hat{F}_2$ be an estimate satisfying $F_2 \leq \hat{F}_2 \leq (1 + 0.01/(2p))F_2$ with probability $1 - n^{-25}$ and is computed using random bits that are independent of the ones used in the above structures.

Let $\bar{\epsilon} = (B/C)^{1/2} = 1/(27p)$. The level-wise thresholds are defined as follows.

$$T_0 = \left(\frac{\hat{F}_2}{B}\right)^{1/2}, \; T_l = \left(\frac{1}{2\alpha}\right)^{l/2} T_0, \quad l \in [L-1], \text{ and}$$
$$Q_l = T_l - \bar{\epsilon} T_l, \; l \in \{0\} \cup [L-1], \quad Q_L = 1/2 \; . \tag{3}$$

Let $\hat{f}_{il}$ be the estimate for f_i obtained from level l using HH_l. For $l \in \{0\} \cup [L-1]$, we say that i is "discovered" at level l, or that $l_d(i) = l$, if l is the smallest level such that $|\hat{f}_{il}| \geq Q_l$. Define $\hat{f}_i = \hat{f}_{i,l_d(i)}$. $l_d(i)$ is set to L iff $i \in \mathcal{S}_L$ and i has not been discovered at any earlier level.

Items are placed into sample groups, denoted by $\bar{G}_l$, for $l \in \{0\} \cup [L]$, as follows. An item is placed into the sampled group $\bar{G}_l$ if the following holds.

1. If i is discovered at level l and $|\hat{f}_{il}| \geq T_l$, then, i is included in $\bar{G}_l$.
2. If i is discovered at level $l-1$ but $|\hat{f}_{i,l-1}| < T_{l-1}$ and the flip of an unbiased coin K_i turns up *heads*.

An item i is placed in G_0 if $|\hat{f}_{i0}| \geq T_0$. In other words, the sample groups are defined as follows.

$$\bar{G}_0 = \{i : |\hat{f}_i| \geq T_0\},$$
$$\bar{G}_l = \{i : (l_d(i) = l \text{ and } |\hat{f}_i| \geq T_l) \text{ or } (l_d(i) = l-1 \text{ and } |\hat{f}_i| < T_{l-1} \text{ and } K_i = 1)\}, \; l = 1, 2, \ldots, L-1,$$
$$\bar{G}_L = \{i : l_d(i) = L \text{ or } (l_d(i) = L-1 \text{ and } |\hat{f}_i| < T_{L-1} \text{ and } K_i = 1)\} \; .$$

We refer to an item as being *sampled* if it belongs to a sample group. From the construction above, it follows that (1) only an item that is discovered may be sampled, and (2) if $i \in [n]$ is discovered at level l, then, i may belong to sampled group $\bar{G}_l$ or to the sampled group $\bar{G}_{l+1}$, or to neither (and hence to no sampled group). That is, there is a possibility that discovered items are not sampled (this happens when $Q_l \leq \hat{f}_{il} < T_l$ and $K_i = 0$ (tails)).

The NOCOLLISION event

Let $\widehat{\text{TOPK}}_l(C_l)$ be the set of the top-C_l elements in terms of the estimates $|\hat{f}_{il}|$ at level l. For $l \in \{0\} \cup [L]$, NOCOLL_l is said to hold if for each $i \in \widehat{\text{TOPK}}_l(C_l)$, there exists a set $R_l(i) \subset [2s]$ of indices of hash tables of the structure TPEST_l such that $|R_l(i)| \geq s$ and that i does not collide with any other item of $\widehat{\text{TOPK}}_l(C_l)$ in the buckets $h_{lq}(i)$, for $q \in R_l(i)$. More precisely,

$$\text{NOCOLL}_l \equiv \forall i \in \widehat{\text{TOPK}}_l(C_l), \exists R_l(i) \subset [2s] \, (|R_l(i)| \geq s \text{ and}$$
$$\forall q \in R_l(i), \forall j \in \widehat{\text{TOPK}}_l(C_l) \setminus \{i\} \;\; h_{lq}(i) \neq h_{lq}(j)\Big) \; . \tag{4}$$

The event NOCOLL is defined as

$$\text{NOCOLL} \equiv \wedge_{l=0}^{L} \text{NOCOLL}_l \ .$$

The analysis shows NOCOLL to be a very high probability event, however, if NOCOLL fails, then, the estimate for F_p returned is 0.

The estimator $\hat{F}_p$

Assume that the event NOCOLL holds, otherwise, $\hat{F}_p$ is set to 0. For each item i that is discovered at level $l_d(i) < L$ and is sampled into sampled group at level $l_s(i)$, the averaged Taylor polynomial estimator is used to obtain an estimate of $|f_i|^p$ using the structure $\text{TPEST}_{l_d(i)}$ at level $l_d(i)$ and scaled by factor of $2^{l_s(i)}$ to compensate for sampling. If $l_d(i) = l_s(i) = L$, then the simpler estimator $|\hat{f}_i|^p$ is used instead and the resulting estimate is scaled by 2^L.

The parameter λ used in the Taylor polynomial estimator for estimating $|f_i|^p$ is set to $|\hat{f}_i| = |\hat{f}_{i,l_d(i)}|$. Let $l = l_d(i)$. By NOCOLL, let $R_l(i) = \{t_1, t_2, \ldots, t_s\} \subset [2s]$. Let X_{ijl} be the (standard) estimate for $|f_i|$ obtained from table T_{lj}, that is,

$$X_{ijl} = T_{lj}[h_{lj}(i)] \cdot \xi_{lj}(i) \cdot \text{sgn}(\hat{f}_i), \qquad \text{for } j \in R_l(i).$$

The estimator $\bar{\vartheta}_i$ is defined as

$$\bar{\vartheta}_i = \bar{\vartheta}(t^p, |\hat{f}_i|, k, s, Y, \{\pi_j\}_{j\in Y}, \{X_{ijl}\}_{j\in R_l(i)})$$

where, Y is a code satisfying Corollary 5 and $\{\pi_j\}_{j\in Y}\}$ is a family of independently and randomly chosen permutations from $[k] \to [k]$. The parameters k and s are given in Figure 2. The estimator $\hat{F}_p$ for F_p is defined below.

$$\hat{F}_p = \sum_{l=0}^{L} \sum_{i\in\bar{G}_l, l_d(i)<L} 2^l \cdot \bar{\vartheta}_i + \sum_{i\in\bar{G}_L, l_d(i)=L} 2^L \cdot |\hat{f}_i|^p \ . \tag{5}$$

4 Analysis

In this section, we analyze the Geometric-Hss algorithm.

4.1 The event $\mathcal{G}$

Let $F_2^{\text{res}}(k, l)$ denote the (random) k-residual second moment of the frequency vector corresponding to $\mathcal{S}_l$. The analysis is conditioned on the conjunction of a set of events denoted by $\mathcal{G}$, as defined in Figure 3.

The events comprising $\mathcal{G}$ are as follows. GOODF_2 is the event that $\hat{F}_2$ is an $1 + O(1/p)$-factor approximation of F_2. The event GOODEST states that for all $i \in [n]$ and levels $l \in \{0\} \cup [L]$, the frequency estimation errors incurred by the HH_l structure remains within the high-probability error bound for the CountSketch algorithm [12] given by Eqn. (1). However, the bounds in GOODEST have to be expressed in terms of $F_2^{\text{res}}(2C_l, l)$, which are themselves random variables. The event SMALLRES gives some control on this random variable by giving an upper bound on $F_2^{\text{res}}(2C_l, l)$ as $\frac{1.5F_2^{\text{res}}((2\alpha)^l C))}{2^{l-1}}$. The event ACCUEST holds if the frequency estimation for an item i at a certain

(1) GOODF$_2$ $\equiv F_2 \le \hat{F}_2 \le \left(1 + \frac{0.001}{2p}\right) F_2,$

(2) NOCOLL defined in (4)

(3) GOODEST $\equiv \forall l : 0 \le l \le L,\ \forall i \in [n],\ \ |\hat{f}_{il} - f_i| \le \left(\frac{F_2^{\text{res}}(2C_l, l)}{C_l}\right)^{1/2}$

(4) SMALLRES $\equiv \forall l : 0 \le l \le L,\ F_2^{\text{res}}(2C_l, l) \le \frac{1.5F_2^{\text{res}}\left((2\alpha)^l C\right)}{2^{l-1}}$

(5) ACCUEST $\equiv \forall l : 0 \le l \le L,\ \forall i \in [n],\ |\hat{f}_{il} - f_i| \le \left(\frac{F_2^{\text{res}}\left((2\alpha)^l C\right)}{2(2\alpha)^l C}\right)^{1/2}$

(6) GOODFINALLEVEL $\equiv \forall i \in \mathcal{S}_L, \hat{f}_{iL} = f_i$

(7) SMALLHH $\equiv \forall l : 0 \le l \le L, \{i : |\hat{f}_{il}| \ge Q_l\} \subset \overline{\text{TOPK}}(C_l).$

Figure 3: $\mathcal{G}$ is the conjunction of these 7 events

level l has an additive accuracy of $\frac{F_2^{\text{res}}\left((2\alpha)^l C)\right)}{(2\alpha)^l C}$. The bounds given by ACCUEST are non-random functions of l. An item i is *classified as a heavy-hitter at level l* if $\hat{f}_{il} \ge Q_l$, that is, its estimate obtained from the HH_l structure exceeds the threshold Q_l. The event SMALLHH is said to hold if at each level, each heavy-hitter item at that level is among those with the top-C_l absolute estimated frequencies at that level. The NOCOLLISION event is used only by the TPEST family of structures at each level, and ensures that each heavy-hitter remains isolated from all the other heavy-hitters of that level in at least half (s) of the tables of the TPEST structure at that level.

Lemma 7 shows that $\mathcal{G}$ holds except with inverse polynomial probability.

Lemma 7. *For the choice of parameters in Figure 2, $\mathcal{G}$ holds with probability $1 - O(n^{-24})$.*

4.2 Grouping items by frequencies

Items are divided into groups based upon frequency ranges, as follows.

$$\begin{aligned} G_0 &= \{i : |f_i| \ge T_0\} \\ G_l &= \{i : T_l \le |f_i| < T_{l-1}\}, l = 1, 2, \ldots, L-1 \\ G_L &= \{i : 1 \le |f_i| < T_{L-1}\}\ . \end{aligned}$$

Note that this grouping is for purposes of analysis, since the true frequencies are unknown to the algorithm. Since estimated frequencies may have errors, it is possible that the sampling algorithm samples an item i into the sampled group $\bar{G}_l$, although, the item does not belong to the group G_l. It will be useful to understand the conditions under which such errors do not occur, and the conditions under which such errors may occur and their extent.

Each group is further partitioned into subsets defined by frequency ranges, namely, lmargin(G_l),

$\text{mid}(G_l)$ and $\text{rmargin}(G_l)$.

$$\begin{aligned}
\text{lmargin}(G_l) &= \{i : T_l \le |f_i| < T_l(1+\bar{\epsilon})\}, \quad l = 0, \ldots, L-1, \\
\text{rmargin}(G_l) &= \{i : T_{l-1}(1-2\bar{\epsilon}) \le |f_i| < T_{l-1}\}, \quad l \in [L] \\
\text{mid}(G_l) &= \{i : T_l + T_l\bar{\epsilon} \le |f_i| < T_{l-1} - 2T_{l-1}\bar{\epsilon}\}, \quad l \in [L-1], \\
\text{mid}(G_0) &= \{i : |f_i| \ge T_0(1+\bar{\epsilon})\} \\
\text{mid}(G_L) &= \{1 \le |f_i| < T_{L-1}(1-2\bar{\epsilon})\} \ .
\end{aligned}$$

G_0 and G_L have no $\text{rmargin}(G_0)$ and $\text{lmargin}(G_L)$ defined, respectively. These definitions are similar (though not identical) to the Hss algorithm [15]. The ratio $\frac{T_{l-1}}{T_l} = (2\alpha)^{1/2}$, for $l = 1, 2, \ldots, L-1$. The last group G_L has frequency range is $[1, T_{L-1})$ and the frequency ratio $T_{L-1}/1$ can be large.

4.3 Properties of the sampling scheme

In the remainder of this paper, we assume that $c > 23$ is a constant satisfying $\mathsf{Pr}[\neg\mathcal{G}] / \mathsf{Pr}[\mathcal{G}] \le n^{-c}$.

Basic Property

Lemma 8 presents the basic property of the sampling scheme.

Lemma 8. *Let $i \in G_l$.*

1. *Let $i \in mid(G_l)$. Then,*
$$\left|2^l \mathsf{Pr}\left[i \in \bar{G}_l \mid \mathcal{G}\right] - 1\right| \le 2^l n^{-c} \ .$$
Further, conditional on $\mathcal{G}$, (i) $i \in \bar{G}_l$ iff $i \in \mathcal{S}_l$, and, (ii) i may not belong to any $\bar{G}_{l'}$, for $l' \ne l$, that is, (i) $\mathsf{Pr}\left[i \in \bar{G}_l \mid \mathcal{G}\right] = \mathsf{Pr}[i \in \mathcal{S}_l \mid \mathcal{G}] = 2^l \pm n^{-c}$, and, (ii) $\mathsf{Pr}\left[i \in \cup_{l' \ne l} \bar{G}_{l'} \mid \mathcal{G}\right] = 0$.

2. *Let $i \in lmargin(G_l)$. Then*
$$\left|2^{l+1} \mathsf{Pr}\left[i \in \bar{G}_{l+1} \mid \mathcal{G}\right] + 2^l \mathsf{Pr}\left[i \in \bar{G}_l \mid \mathcal{G}\right] - 1\right| \le 2^l n^{-c} \ .$$
Further, conditional on $\mathcal{G}$, i may belong to either $\bar{G}_l$ or $\bar{G}_{l+1}$, but not to any other sampled group, that is, $\mathsf{Pr}\left[i \in \cup_{l' \notin \{l, l+1\}} \bar{G}_{l'} \mid \mathcal{G}\right] = 0$.

3. *If $i \in rmargin(G_l)$, then*
$$\left|2^l \mathsf{Pr}\left[i \in \bar{G}_l \mid \mathcal{G}\right] + 2^{l-1} \mathsf{Pr}\left[i \in \bar{G}_{l-1} \mid \mathcal{G}\right] - 1\right| \le O(2^l n^{-c}) \ .$$
Further, conditional on $\mathcal{G}$, i can belong to either $\bar{G}_{l-1}$ or $\bar{G}_l$ and not to any other sampled group, that is, $\mathsf{Pr}\left[i \in \cup_{l' \notin \{l-1, l\}} \bar{G}_{l'} \mid \mathcal{G}\right] = 0$.

Lemma 8 is essentially true (with minor changes) for the Hss method [6, 15], although the Hss analysis used full-independence of hash functions whereas here we work with limited independence. A straightforward corollary of Lemma 8 is the following.

Corollary 9. *Let $i \in G_l$. Then,*

$$\sum_{l'=0}^{L} 2^{l'} \mathsf{Pr}\left[i \in \bar{G}_{l'} \mid \mathcal{G}\right] = \sum_{l' \in \{0,1,\ldots,L\} \cap \{l-1, l, l+1\}} \mathsf{Pr}\left[i \in \bar{G}_{l'} \mid \mathcal{G}\right] = 1 \pm 2^{l+1} n^{-c} \ .$$

Approximate pair-wise independence property

Lemma 10 essentially repeats the results of Lemma 8, conditional upon the event that another item maps to a substream at some level l. This property is a step towards proving an approximate pair-wise independence property in the following section.

Lemma 10. *Let $i, j \in [n]$, $i \neq j$ and $j \in G_r$.*

1. *Let $j \in mid(G_r)$. Then*

$$\left|2^r \mathsf{Pr}\left[j \in \bar{G}_r \mid i \in \mathcal{S}_l, \mathcal{G}\right] - 1\right| \leq 2^r n^{-c} \ .$$

   Further, for any $r \neq r'$, $\mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_l, \mathcal{G}\right] = 0$.

2. *Let $j \in lmargin(G_r)$. Then,*

$$\left|2^{r+1}\mathsf{Pr}\left[j \in \bar{G}_{r+1} \mid i \in \mathcal{S}_l, \mathcal{G}\right] + 2^r \mathsf{Pr}\left[j \in \bar{G}_r \mid i \in \mathcal{S}_l, \mathcal{G}\right] - 1\right| \leq 2^{r+1} n^{-c} \ .$$

   Further, for any $r' \notin \{r, r+1\}$, $\mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_l, \mathcal{G}\right] = 0$.

3. *If $j \in rmargin(G_r)$, then*

$$\left|2^r \mathsf{Pr}\left[j \in \bar{G}_r \mid i \in \mathcal{S}_l, \mathcal{G}\right] + 2^{r-1}\mathsf{Pr}\left[j \in \bar{G}_{r-1} \mid i \in \mathcal{S}_l, \mathcal{G}\right] - 1\right| \leq 2^{r+1} n^{-c} \ .$$

   Further, for any $r' \notin \{r-1, r\}$, $\mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_l, \mathcal{G}\right] = 0$.

Corollary 11. *Let $i, j \in [n]$, $i \neq j$ and $j \in G_r$. Then,*

$$\left|\sum_{r'=0}^{L} 2^{r'} \mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_l, \mathcal{G}\right] - 1\right| \leq O(2^r n^{-c}) \ .$$

We can now prove an approximate pair-wise independence property.

Lemma 12. *For $i \in G_l$, $j \in G_m$ and i, j distinct,*

$$\left|\sum_{r,r'=0}^{L} 2^{r+r'} \mathsf{Pr}\left[i \in \bar{G}_r, j \in \bar{G}_{r'} \mid \mathcal{G}\right] - 1\right| \leq O((2^l + 2^m) n^{-c}) \ .$$

4.4 Application of Taylor Polynomial Estimator

Let $i \in \bar{G}_{l'}$ for some $l' \in \{0\} \cup [L-1]$. Then, i has been discovered at a level $l_d(i) = l$ (say). The algorithm estimates $|f_i|^p$ from the TPEST structure at the discovery level l using the estimator

$$\bar{\vartheta}_i = \bar{\vartheta}(\psi(t) = t^p, |\hat{f}_i|, k, s, Y, \{\pi_y\}_{y \in Y}, \{X_{ijl}\}_{j \in R_l(i)}) \ .$$

By construction, $\hat{f}_i$ is defined as $\hat{f}_{il}$ and for any $j \in R_l(i)$, $\sigma_{ijl} = (\mathsf{Var}\left[X_{ijl}\right])^{1/2}$ and $\eta_{ijl} = (\sigma_{il}^2 + (|f_i| - |\hat{f}_{il}|)^2$. We first show that the premises of Corollary 2 and Lemma 6 are satisfied so that we can use their implications.

Lemma 13. *Assume the parameter values listed in Figure 2 and that $\mathcal{G}$ holds. Suppose $l_d(i) = l$ for some $l \in \{0\} \cup [L-1]$. Then the following properties hold.*

a) $|\hat{f}_{il} - f_i| \leq |f_i|/(26p)$,

b) $\mathbb{E}\left[X_{ijl} \mid l_d(i) = l, |\hat{f}_{il}| > Q_l, j \in R_l(i), \mathcal{G}\right] = |f_i|$,

c) $|f_i| \geq 15p\eta_{ijl_d(i)}$, *for* $j \in R_{l_d(i)}(i)$,

d) $\eta^2_{ijl_d(i)} \leq 2.7(\bar{\epsilon}T_l)^2$, *for* $j \in R_{l_d(i)}(i)$,

e) $|\hat{f}_{il} - f_i| \leq |\hat{f}_i|/(26p)$,

f) $|\hat{f}_i|/\eta_{ijl_d(i)} \geq 16p$, *for* $j \in R_{l_d(i)}(i)$,

g) if $l_d(i) = L$, *then,* $\hat{f}_i = f_i$ *and* $\eta_{iL} = 0$.

For $i, k \in \mathcal{S}_l$, $j \in [2s]$, let $u_{ikjl} = 1$ iff $h_{lj}(i) = h_{lj}(k)$ and 0 otherwise.

Lemma 14. *Assume the parameters in Figure 2 and let* $p \geq 2$. *Suppose* $i \in \bar{G}_l$, *for some* $l \in \{0\} \cup [L-1]$. *Then,*

$$\left|\mathbb{E}\left[\bar{\vartheta}_i \mid \mathcal{G}\right] - |f_i|^p\right| \leq n^{-4000p}|f_i|^p \ .$$

Further if p *is integral, then,* $\mathbb{E}\left[\bar{\vartheta}_i \mid \mathcal{G}\right] = |f_i|^p$.

We denote by $\bar{\xi}$ the set of random bits defining the family of Rademacher random variables used by the TPEST structures, that is, the set of random bits that defines the family $\{\xi_{lj}(i) \mid i \in [n], j \in [2s], l \in \{0\} \cup [L]\}$. Lemma 15 shows that the event NOCOLL implies that the Taylor polynomial estimators are pair-wise uncorrelated.

Lemma 15. *Suppose* $i \in \bar{G}_r$ *and* $i' \in \bar{G}_{r'}$. *Then,*

$$\mathbb{E}_{\bar{\xi}}\left[\bar{\vartheta}_i\bar{\vartheta}_{i'} \mid \hat{f}_i, \hat{f}_{i'}, \mathcal{G}\right] = \mathbb{E}_{\bar{\xi}}\left[\bar{\vartheta}_i \mid \hat{f}_i, \mathcal{G}\right]\mathbb{E}_{\bar{\xi}}\left[\bar{\vartheta}_{i'} \mid \hat{f}_{i'}, \mathcal{G}\right] \ .$$

4.5 Expectation and Variance of $\hat{F}_p$ Estimator.

For uniformity of notation, let $\bar{\vartheta}_i$ denote $|\hat{f}_i|$ when $l_d(i) = L$ and otherwise, let its meaning be unchanged. Let z_{il} be an indicator variable that is 1 if $i \in \bar{G}_l$ and 0 otherwise. Since an item may be sampled into at most one group, $\sum_{l\in[L]} z_{il} \in \{0, 1\}$. Using the extended definition of $\bar{\vartheta}_i$ mentioned above, we can write $\hat{F}_p$ as,

$$\begin{aligned}
\hat{F}_p &= \sum_{l=0}^{L}\sum_{i\in\bar{G}_l} 2^l\bar{\vartheta}_i \\
&= \sum_{i\in[n]}\sum_{l=0}^{L} z_{il}\cdot 2^l\cdot\bar{\vartheta}_i \\
&= \sum_{i\in[n]} Y_i \qquad (6)
\end{aligned}$$

where,

$$Y_i = \sum_{l'=0}^{L-1} 2^{l'} z_{il'}\bar{\vartheta}_i \ . \qquad (7)$$

Lemma 16 shows that $\hat{F}_p$ is almost an unbiased estimator for F_p. This follows from Lemma 14.

Lemma 16. $\mathbb{E}\big[\hat{F}_p \mid \mathcal{G}\big] = F_p(1 \pm O(n^{-c+1}))$.

We will use the following facts that are easily proved (see Appendix).

$$\begin{aligned} F_2 &\leq n^{1-2/p} F_p^{2/p}, && p \geq 2, \\ F_{2p-2} &\leq F_p^{2-2/p}, && p \geq 2. \end{aligned} \tag{8}$$

Lemma 17. *Let* $B = Kn^{1-2/p}\epsilon^{-2}/\log(n)$ *and* $C = (27p)^2 B$. *Then,*

$$\mathsf{Var}\left[Y_i \mid \mathcal{G}\right] \leq \begin{cases} \dfrac{\epsilon^2 |f_i|^{2p-2} F_p^{2/p}}{(5)(10)^4 K} & \text{if } i \in \mathit{mid}(G_0) \\ 2^{l+1}(1.002)|f_i|^{2p} & \text{if } i \in \mathit{lmargin}(G_0) \cup_{l=1}^{L} G_l \end{cases}$$

Lemma 18 builds on the approximate pair-wise independence of the sampling scheme (Lemma 12) and the pair-wise uncorrelated property of the $\bar{\vartheta}_i$ estimators (Lemma 15) to show that the $\mathsf{Cov}\left(Y_i, Y_j\right)$, for $i \neq j$ is very small.

Lemma 18. *Let* $i \neq j$. *Then,*

$$\left|\mathsf{Cov}\left(Y_i, Y_j \mid \mathcal{G}\right)\right| \leq O(n^{-c+1})|f_i|^p |f_j|^p \ .$$

Lemma 19 gives a bound on the variance of the $\hat{F}_p$ estimator.

Lemma 19.

$$\mathsf{Var}\big[\hat{F}_p \mid \mathcal{G}\big] \leq \frac{\epsilon^2 F_p^2}{50} \ .$$

Putting things together

Theorem 20 states the space bound for the algorithm and the update time.

Theorem 20. *For each fixed* $p > 2$ *and* $0 < \epsilon \leq 1$, *there exists an algorithm in the general update data stream model that returns* $\hat{F}_p$ *satisfying* $\left|\hat{F}_p - F_p\right| < \epsilon F_p$ *with probability* $3/4$. *The algorithm uses space* $O(n^{1-2/p}\epsilon^{-2} + n^{1-2/p}\epsilon^{-4/p}\log(n))$ *words of size* $O(\log(nmM))$ *bits. The time taken to process each stream update is* $O(\log^2 n)$.

Acknowledgement

The author thanks Venugopal G. Reddy for correcting an error in the analysis.

References

[1] Noga Alon, Yossi Matias, and Mario Szegedy. "The space complexity of approximating frequency moments". *Journal of Computer Systems and Sciences*, 58(1):137–147, 1998. Preliminary version appeared in Proceedings of *ACM* Symposium on Theory of Computing (*STOC*) 1996, pp. 1-10.

[2] Alexander Andoni, Robert Krauthgamer, and Krzysztof Onak. "Streaming Algorithms via Precision Sampling". In *Proceedings of IEEE Foundations of Computer Science (FOCS)*, 2011. A version appears in arXiv:1011.1263v1 [cs.DS] November 2010.

[3] Alexandr Andoni, Huy L. Nguyen, Yury Polyanskiy, and Yihong Wu. "Tight Lower Bound for Linear Sketches of Moments". In *Proceedings of International Conference on Automata, Languages and Programming, (ICALP)*, July 2013. Version published as arXiv:1306.6295, June 2013.

[4] Kanh Do Ba, Piotr Indyk, Eric Price, and David Woodruff. "Lower bounds for sparse recovery". In *Proceedings of ACM Symposium on Discrete Algorithms (SODA)*, 2008.

[5] Z. Bar-Yossef, T.S. Jayram, R. Kumar, and D. Sivakumar. "An information statistics approach to data stream and communication complexity". In *Proceedings of ACM Symposium on Theory of Computing STOC*, pages 209–218, 2002.

[6] L. Bhuvanagiri, S. Ganguly, D. Kesh, and C. Saha. "Simpler algorithm for estimating frequency moments of data streams". In *Proceedings of ACM Symposium on Discrete Algorithms (SODA)*, pages 708–713, 2006.

[7] Vladimir Braverman, Jonathan Katzman, Charles Seidell, and Gregory Vorsanger. "Approximating Large Frequency Moments with $O(n^{1-2/k})$ Bits". In *Proceedings of International Workshop on Randomization and Computation (RANDOM)*, 2014. Published earlier as arXiv:1401.1763, January 2014.

[8] Vladimir Braverman and Rafail Ostrovsky. "Recursive Sketching For Frequency Moments". arXiv:1011.2571v1 [cs.DS], November 2010.

[9] Emmanuel Candès, Justin Romberg, and Terence Tao. "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information". *IEEE Trans. Inf. Theory*, 52(2):489–509, February 2006.

[10] Nicolò Cesa-Bianchi, Shai Shalev Shwartz, and Ohad Shamir. "Online Learning of Noisy Data with Kernels". In *Proceedings of ACM International Conference on Learning Theory (COLT)*, 2010.

[11] A. Chakrabarti, S. Khot, and X. Sun. "Near-Optimal Lower Bounds on the Multi-Party Communication Complexity of Set Disjointness". In *Proceedings of International Conference on Computational Complexity (CCC)*, 2003.

[12] Moses Charikar, Kevin Chen, and Martin Farach-Colton. "Finding frequent items in data streams". *Theoretical Computer Science*, 312(1):3–15, 2004. Preliminary version appeared in Proceedings of ICALP 2002, pages 693-703.

[13] Graham Cormode and S. Muthukrishnan. "Combinatorial Algorithms for Compressed Sensing". In *Proceedings of International Colloquium on Structural Information & Communication Complexity, (SIROCCO)*, 2006.

[14] David L. Donoho. "Compressed Sensing". *IEEE Trans. Inf. Theory*, 52(4):1289–1306, April 2006.

[15] S. Ganguly and L. Bhuvanagiri. "Hierarchical Sampling from Sketches: Estimating Functions over Data Streams". *Algorithmica*, 53:549–582, 2009.

[16] S. Ganguly, D. Kesh, and C. Saha. "Practical Algorithms for Tracking Database Join Sizes". In *Proceedings of Foundations of Software Technoogy and Theoretical Computer Science (FSTTCS)*, pages 294–305, Hyderabad, India, December 2005.

[17] Sumit Ganguly. "A Lower Bound for Estimating High Moments of a Data Stream". arXiv:1201.0253, December 2011.

[18] Sumit Ganguly. "Precision vs. Confidence Tradeoffs for ℓ_2-Based Frequency Estimation in Data Streams". In *Proceedings of International Symposium on Algorithms, Automata and Computation (ISAAC), LNCS Vol. 7676*, pages 64–74, 2012.

[19] Piotr Indyk. Stable distributions, pseudorandom generators, embeddings, and data stream computation. *J. ACM*, 53(3):307–323, 2006. Preliminary Version appeared in Proceedings of IEEE FOCS 2000, pages 189-197.

[20] Piotr Indyk and David Woodruff. "Optimal Approximations of the Frequency Moments". In *Proceedings of ACM Symposium on Theory of Computing STOC*, pages 202–298, Baltimore, Maryland, USA, June 2005.

[21] T.S. Jayram and David Woodruff. "Optimal Bounds for Johnson-Lindenstrauss Transforms and Streaming Problems with Low Error". In *Proceedings of ACM Symposium on Discrete Algorithms (SODA)*, 2011.

[22] Hossein Jowhari, Mert Săglam, and Gábor Tardos. "Tight Bounds for Lp Samplers, Finding Duplicates in Streams, and Related Problems". In *Proceedings of ACM International Symposium on Principles of Database Systems (PODS)*, 2011.

[23] Yi Li and David Woodruff. "A Tight Lower Bound for High Frequency Moment Estimation with Small Error". In *Proceedings of International Workshop on Randomization and Computation (RANDOM)*, 2013.

[24] Morteza Monemizadeh and David Woodruff. "1-pass relative-error l_p-sampling with applications". In *Proceedings of ACM Symposium on Discrete Algorithms (SODA)*, 2010.

[25] N. Nisan. "Pseudo-Random Generators for Space Bounded Computation". In *Proceedings of ACM Symposium on Theory of Computing STOC*, pages 204–212, May 1990.

[26] Eric Price and David Woodruff. "$(1+\epsilon)$-approximate Sparse Recovery". In *Proceedings of IEEE Foundations of Computer Science (FOCS)*, 2011.

[27] Oren Patashnik Ronald L. Graham, Donald E. Knuth. *"Concrete Mathematics A Foundation for Computer Science"*. Addison-Wesley, 1994.

[28] J. Schmidt, A. Siegel, and A. Srinivasan. "Chernoff-Hoeffding Bounds with Applications for Limited Independence". In *Proceedings of ACM Symposium on Discrete Algorithms (SODA)*, pages 331–340, 1993.

[29] R. Singh. "Existence of unbiased estimates". *Sankhya: The Indian Journal of Statistics*, 26(1):93–96, 1964.

[30] M. Thorup and Y. Zhang. "Tabulation based 4-universal hashing with applications to second moment estimation". In *Proceedings of ACM Symposium on Discrete Algorithms (SODA)*, pages 615–624, New Orleans, Louisiana, USA, January 2004.

[31] David P. Woodruff. "Optimal space lower bounds for all frequency moments". In *Proceedings of ACM Symposium on Discrete Algorithms (SODA)*, pages 167–175, 2004.

[32] David P. Woodruff and Qin Zhang. "Tight Bounds for Distributed Functional Monitoring". In *Proceedings of ACM Symposium on Theory of Computing STOC*, 2012.

A Proofs for the Taylor Polynomial estimator

Fact 21. *Let $k > p \geq 0$. Then, $\left|\binom{p}{k}\right| \leq \left(\frac{p}{k}\right)^{\lfloor p \rfloor + 1}$. In particular, if $p \in \mathbb{Z}^+$, then, $\binom{p}{k} = 0$.*

Proof. The second statement is obvious, since for $k > p \geq 0$ and p integral, $p^{\underline{k}} = 0$. Otherwise, for non-integral p, using the absorption identity $\lfloor p \rfloor + 1$ times, gives

$$\binom{p}{k} = \left(\frac{p^{\underline{\lfloor p \rfloor + 1}}}{k^{\underline{\lfloor p \rfloor + 1}}}\right)\binom{p-\lfloor p \rfloor - 1}{k-\lfloor p \rfloor - 1} = \left(\frac{p^{\underline{\lfloor p \rfloor + 1}}}{k^{\underline{\lfloor p \rfloor + 1}}}\right)(-1)^k\binom{k-p-1}{k-\lfloor p \rfloor - 1}$$

Now, for $0 \leq j \leq \lfloor p \rfloor$, $\frac{p-j}{k-j} \leq \frac{p}{k}$, since $p < k$. Therefore, $\frac{p^{\underline{\lfloor p \rfloor + 1}}}{k^{\underline{\lfloor p \rfloor + 1}}} \leq \left(\frac{p}{k}\right)^{\lfloor p \rfloor + 1}$. Similarly, $\binom{k-p-1}{k-\lfloor p \rfloor - 1} \leq \left(\frac{k-p-1}{k-\lfloor p \rfloor - 1}\right)^{k-\lfloor p \rfloor - 1} < 1$. Taking absolute values, $\left|\binom{p}{k}\right| \leq \left(\frac{p}{k}\right)^k$. □

Proof of Lemma 1. Fix ψ, λ and k and let $\vartheta = \vartheta(\psi, \lambda, k, X_1, \ldots, X_k)$. Using linearity of expectation and independence of X_i's we have,

$$\mathbb{E}[\vartheta] = \mathbb{E}\left[\sum_{j=0}^{k}\gamma_j(\lambda)\prod_{v=1}^{j}(X_v - \lambda)\right] = \sum_{j=0}^{k}\gamma_j(\lambda)\prod_{v=1}^{j}(\mu - \lambda) = \psi(\lambda + \mu - \lambda) - \gamma_{k+1}(\lambda')(\mu - \lambda)^{k+1}$$

for some $\lambda' \in (\mu, \lambda)$ by the Taylor series expansion of $\psi(\mu) = \psi(\lambda + (\mu - \lambda))$ around λ. The Taylor series expansion of $\psi(\mu)$ around λ exists since ψ is analytic in the interval $[\mu, \lambda]$. Therefore,

$$|\mathbb{E}[\vartheta] - \psi(\mu)| \leq |\gamma_{k+1}(\lambda')||\mu - \lambda|^{k+1}$$

proving part (i) of the lemma.
For $j = 0, 1, \ldots, k$, let

$$P_j = \prod_{l=1}^{j}(X_l - \lambda)$$

(which implies that $P_0 = 1$). Then,

$$\vartheta = \sum_{j=0}^{k}\gamma_j(\lambda)P_j \ .$$

By the independence of the X_l's,

$$\begin{aligned}\mathsf{Var}\left[P_j\right] = \mathsf{Var}\left[\prod_{l=1}^{j}(X_l-\lambda)\right] &= \prod_{l=1}^{j}\mathbb{E}\left[(X_l-\lambda)^2\right] - \prod_{l=1}^{j}(\mathbb{E}\big[X_l-\lambda\big])^2 \\ &= \eta^{2j} - (\mu-\lambda)^{2j} \ .\end{aligned}$$

Further for $1 \leq j < j' \leq k$,

$$\begin{aligned}\mathsf{Cov}\left(P_j, P_{j'}\right) &= \mathsf{Cov}\left(\prod_{l=1}^{j}(X_l-\lambda), \prod_{l=1}^{j'}(X_l-\lambda)\right) \\ &= \mathbb{E}\left[\prod_{l=1}^{j}(X_l-\lambda)\prod_{l=1}^{j'}(X_l-\lambda)\right] - (\mu-\lambda)^{j+j'} \\ &= \prod_{l=1}^{j}\mathbb{E}\left[(X_l-\lambda)^2\right]\prod_{l=j+1}^{j'}\mathbb{E}\left[X_l-\lambda\right] - (\mu-\lambda)^{j+j'} \\ &= \eta^{2j}(\mu-\lambda)^{j'-j} - (\mu-\lambda)^{j+j'} \ .\end{aligned}$$

Thus we have,

$$\begin{aligned}&\mathsf{Var}\left[\vartheta\right] = \sum_{j=0}^{k}(\gamma_j(\lambda))^2\mathsf{Var}\left[P_j\right] + \sum_{j<j'}2\gamma_j(\lambda)\gamma_{j'}(\lambda)\mathsf{Cov}\left(P_j, P_{j'}\right) \\ &= \sum_{j=0}^{k}(\gamma_j(\lambda))^2(\eta^{2j}-(\mu-\lambda)^{2j}) + \sum_{0\leq j<j'\leq k}2\gamma_j(\lambda)\gamma_{j'}(\lambda)(\eta^{2j}(\mu-\lambda)^{j'-j} - (\mu-\lambda)^{j+j'}) \\ &= \sum_{j=1}^{k}(\gamma_j(\lambda))^2(\eta^{2j}-(\mu-\lambda)^{2j}) + \sum_{1\leq j<j'\leq k}2\gamma_j(\lambda)\gamma_{j'}(\lambda)(\eta^{2j}(\mu-\lambda)^{j'-j} - (\mu-\lambda)^{j+j'}) \\ &= \sum_{j=1}^{k}(\gamma_j(\lambda))^2(\eta^{2j}-(\mu-\lambda)^{2j}) + \sum_{1\leq j<j'\leq k}2\gamma_j(\lambda)\gamma_{j'}(\lambda)\prod_{i\in Q^{vv'}}\eta^{2j}(\mu-\lambda(^{j'-j}\left(1-\left(\frac{(\mu-\lambda)^2}{\eta^2}\right)^j\right)\end{aligned} \tag{9}$$

Let $t_j = (\mu-\lambda)^{j'-j}\eta^{2j}\big(1-\big(\frac{(\mu-\lambda)^2}{\eta^2}\big)^{2j}\big)$. Since, $\eta^2 = \sigma^2 + (\mu-\lambda)^2$, we have, $|t_j| \leq |\mu-\lambda|^{j'-j}\eta^{2j} \leq \eta^{j+j'}$. Taking absolute values on both sides of Eqn. (9), we have,

$$\begin{aligned}\mathsf{Var}\left[\vartheta\right] &\leq \sum_{j=1}^{k}\gamma_j^2(\lambda)\eta^{2j} + \sum_{1\leq j<j'\leq k}2|\gamma_j(\lambda)||\gamma_{j'}(\lambda)|\eta^{j+j'} \\ &= \left(\sum_{j=1}^{k}|\gamma_j(\lambda)|\eta^j\right)^2 \ .\end{aligned}$$

□

Proof of Corollary 2. $\lambda \geq \mu(1-\alpha) > 0$ since, $0 \leq \alpha < 1$ and $\mu > 0$. Hence, $\psi(t) = t^p$ is analytic in the interval $[\mu, \lambda]$ (or, $[\lambda, \mu]$ depending on whether $\mu < \lambda$ or $\lambda < \mu$).

Let ϑ abbreviate $\vartheta(\psi(t) = t^p, \lambda, k, \{X_l\}_{l=1}^k)$. Note that for the function $\psi(t) = t^p$, $\gamma_k(w) = \frac{1}{k!}\left(\frac{d^k}{dt^k}t^p\right)\Big|_{t=w} = \binom{p}{k}w^{p-k}$. Applying Lemma 1, there exists $\lambda' \in (\lambda, \mu)$ such that,

$$\begin{aligned}
\left|\mathbb{E}\left[\vartheta\right] - \mu^p\right| &= \left|\gamma_{k+1}(\lambda')\right||\mu-\lambda|^{k+1} = \left|\binom{p}{k+1}\right|\left|\lambda'^{p-k-1}\right||\mu-\lambda|^{k+1} \\
&\leq \left(\frac{p}{k+1}\right)^{\lfloor p\rfloor+1}\mu^{p-k-1}\left(1-\alpha\right)^{p-k-1}\left(\alpha\mu\right)^{k+1}, \quad \text{since, } k+1 > p \text{ and by Fact 21} \\
&= \left(\frac{p}{k+1}\right)^{\lfloor p\rfloor+1}\left(\frac{\alpha}{1-\alpha}\right)^{k+1}(1-\alpha)^p\mu^p \ .
\end{aligned}$$

In particular, if p is integral, then, $\binom{p}{k+1} = 0$ and $\mathbb{E}\left[\vartheta\right] = \mu^p$. □

Proof of Corollary 3. For $\psi(t) = t^p$, $\gamma_v(\lambda) = \binom{p}{v}\lambda^{p-v}$. We also have from the assumptions that $\eta^2 = (\mu-\lambda)^2 + \sigma^2 \leq 2(\frac{\lambda}{25p})^2$, or, $\frac{\eta}{\lambda} \leq \frac{\sqrt{2}}{25}$.

By Lemma 1, part (2),

$$\mathsf{Var}\left[\vartheta\right] \leq \left(\sum_{v=1}^{k}\left|\binom{p}{v}\right|\left|\lambda^{p-v}\eta^v\right|\right)^2 = \lambda^{2p-2}\eta^2\left(\sum_{v=1}^{k}\left|\binom{p}{v}\right|\left|\left(\frac{\eta}{\lambda}\right)^{v-1}\right|\right)^2 \tag{10}$$

The ratio of the $(v+1)$st term in the summation in the *RHS* to the vth term, for $1 \leq v \leq k-1$, is

$$\left|\frac{p-v}{v+1}\right| \cdot \frac{\eta}{\lambda} \leq \frac{(p-1)\sqrt{2}}{2(25p)} < \frac{1}{25\sqrt{2}}$$

Substituting in Eqn. (10) for $\mathsf{Var}\left[\vartheta\right]$ and using $\lambda \leq \mu(1+\frac{1}{25p}) \leq e^{1/(25p)}\mu$, we have,

$$\mathsf{Var}\left[\vartheta\right] \leq \lambda^{2p-2}\eta^2p^2\left(\sum_{v=1}^{k}(25\sqrt{2})^{-(v-1)}\right)^2 \leq (1.08)p^2\mu^{2p-2}\eta^2 \ .$$

□

B Proofs for Averaged Taylor Polynomial Estimator

Proof of Corollary 5. Choosing $q = 8$ and $\epsilon = 3/4$ in Theorem 4 gives a code $Y \subset \{0,1\}^{8k}$ of binary vectors with exactly k 1's and minimum distance $3k/2$. So, $H_q(\epsilon) = 0.9722648\ldots$ and hence, by Theorem 4, $\log|Y| > (1-H_q(\epsilon))k\log 8$ or, $|Y| > 2^{3(1-H_q(\epsilon))k} > 2^{0.08k}$. □

Recall that $Y \subset \{0,1\}^s$ where, $s = 8k$, is a code such that every $y \in Y$ has exactly k 1's, and the minimum Hamming distance between any pair of codewords in Y is at least $3k/2$. Equivalently, y can be written as an ordered sequence $(y_1, y_2, \ldots, y_k)$ where, $1 \leq y_1 < y_2 < \ldots < y_k \leq s$ are the coordinates of the position of 1's in the s-dimensional binary vector y. For example, let $s = 4$ and $k = 2$—then the vector $(1,0,1,0)$ is written as the 2-dimensional ordered sequence $(1,3)$. We will

say that $u \in y$ if u is one of the y_i's in the ordered sequence notation. This notation views the sequence $(1,3)$ above as a set $\{1,3\}$.

Given codewords $y, y' \in Y$, $y \cap y'$ denotes the set of indices that are 1 in both y and y'. Let $\pi : [k] \to [k]$ be a permutation and $y = (y_1, \ldots, y_k)$ be an ordered sequence of size k. Then, $\pi(y)$ denotes the sequence $(y_{\pi(1)}, y_{\pi(2)}, \ldots, y_{\pi(k)})$. The prefix-segment of $\pi(y)$ consisting of its first v entries is $(y_{\pi(1)}, \ldots, y_{\pi(v)})$. Let y, y' be ordered sequences of length k and let π, π' be permutations mapping $[k] \to [k]$. Let $Q^{vv'}_{yy'\pi\pi'}$ denote the set of common indices shared among the first v positions of $\pi(y)$ with the first v' positions of $\pi'(y')$, that is,

$$Q^{vv'}_{yy'\pi\pi'} = \{y_{\pi(1)}, y_{\pi(2)}, \ldots, y_{\pi(v)}\} \cap \{y'_{\pi'(1)}, y'_{\pi'(2)}, \ldots, y'_{\pi'(v')}\} \ .$$

Let $q^{vv'}_{yy'\pi\pi'}$ denote the number of common indices, that is,

$$q^{vv'}_{yy'\pi\pi'} = \left|Q^{vv'}_{yy'\pi\pi'}\right| \ .$$

Given distinct codewords $y, y' \in Y$ and permutations π and π', $Q^{vv'}_{yy'\pi\pi'}$ is abbreviated as $Q^{vv'}$ and $q^{vv'}_{yy'\pi\pi'}$ as $q^{vv'}$.

In the remainder of this section, we will assume that Y is a code of $s = 8k$-dimensional boolean vectors of size exponential in k, as given by Corollary 5. The function for the Taylor polynomial estimator will be $\psi(t) = t^p$. Let ϑ_y abbreviate the estimator $\vartheta_y \equiv \vartheta(\psi(t) = t^p, \lambda, k, s, y, \pi_y, \{X_l\}_{l=1}^s)$, where, λ is some parameter.

B.1 Covariance of $\vartheta_y, \vartheta_{y'}$

Lemma 22. *Let $q = 8$, $k > 1$ and $s = qk$. Let Y be a code satisfying Corollary 5. Let $\{X_1, \ldots, X_s\}$ be a family of independent random variables, each having expectation $\mu > 0$ and variance bounded above by σ^2. Let λ be an estimate for μ satisfying $|\lambda - \mu| \le \min(\mu, \lambda)/(25p)$ and let $\sigma < \min(\mu, \lambda)/(25p)$. Let $\eta = ((\lambda-\mu)^2 + \sigma^2)^{1/2} > 0$. Let $\bar{\vartheta}$ denote $\bar{\vartheta}(t^p, \lambda, k, s, Y, \{\pi_y\}_{y\in Y}, \{X_l\}_{l=1}^s)$ and let ϑ_y denote the estimator $\vartheta_y = \vartheta(t^p, \lambda, k, s, y, \pi_y, \{X_l\}_{l=1}^s)$. Then, for $y, y' \in Y$ and $y \neq y'$,*

$$\mathsf{Cov}\left(\vartheta_y, \vartheta_{y'}\right) = \begin{cases} \displaystyle\sum_{v,v'=1}^{k} \gamma_v(\lambda)\gamma_{v'}(\lambda)(\mu-\lambda)^{v+v'} \mathbb{E}_{\pi_y,\pi_{y'}}\left[\left(\frac{\eta^2}{(\mu-\lambda)^2}\right)^{q^{vv'}} - 1\right] & \text{if } \mu \neq \lambda, \\ \displaystyle = \sum_{v=1}^{k} \gamma_v^2(\lambda)\eta^{2v} \mathsf{Pr}_{\pi_y,\pi_{y'}}\left[q^{vv}_{yy'\pi_y\pi_{y'}} = v\right] & \text{if } \mu = \lambda. \end{cases}$$

Proof of Lemma 22. By definition, $\bar{\vartheta} = \frac{1}{|Y|}\sum_{y\in Y}\vartheta_y$. Fix $y, y' \in Y$, with $y \neq y'$ and let $\pi = \pi_y$ and $\pi' = \pi_{y'}$ abbreviate the random permutations corresponding to y and y'. Let $q^{vv'}_{yy'\pi_y\pi_{y'}}$ be denoted by $q^{vv'}$. Now,

$$\mathbb{E}\left[\vartheta_y\right]\mathbb{E}\left[\vartheta_{y'}\right] = \left(\sum_{v=0}^{k}\gamma_v(\lambda)(\mu-\lambda)^v\right)^2 = \sum_{v=0}^{k}\sum_{v'=0}^{k}\gamma_v(\lambda)\gamma_{v'}(\lambda)(\mu-\lambda)^{v+v'} \ .$$

Further, from the definition of ϑ_y and $\vartheta_{y'}$, and by linearity of expectation,

$$\begin{aligned}\mathbb{E}\left[\vartheta_y\vartheta_{y'}\right] &= \mathbb{E}\left[\left(\sum_{v=0}^{k}\gamma_v(\lambda)\prod_{l=1}^{v}(X_{y_{\pi(l)}}-\lambda)\right)\left(\sum_{v'=0}^{k}\gamma_{v'}(\lambda)\prod_{m=1}^{v'}(X_{y'_{\pi'(m)}}-\lambda)\right)\right]\\ &= \sum_{v,v'=0}^{k}\gamma_v(\lambda)\gamma_{v'}(\lambda)\mathbb{E}\left[\prod_{l=1}^{v}(X_{y_{\pi(l)}}-\lambda)\prod_{m=1}^{v'}(X_{y'_{\pi'(m)}}-\lambda)\right]\end{aligned}$$

Fix π, π'. There are $q^{vv'} = q^{vv'}_{yy'\pi_y\pi_{y'}}$ indices that are common among the first v positions of $\pi_y(y)$ and the first v' positions of $\pi_{y'}(y')$. This set of common indices is given by $Q^{vv'} = Q^{vv'}_{yy'\pi_y\pi_{y'}} = \{y_{\pi(1)},\ldots,y_{\pi(v)}\}\cap\{y'_{\pi'(1)},\ldots,y'_{\pi'(v)}\}$. Also, let $U^{vv'} = U^{vv'}_{yy'\pi_y\pi_{y'}}$ denote the union $\{y_{\pi(1)},\ldots,y_{\pi(v)}\}\cup\{y'_{\pi'(1)},\ldots,y'_{\pi'(v)}\}$. Hence we have,

$$\prod_{l=1}^{v}(X_{y_{\pi(l)}}-\lambda)\prod_{m=1}^{v'}(X_{y'_{\pi'(m)}}-\lambda) = \prod_{i\in Q^{vv'}}(X_i-\lambda)^2\prod_{i\in U^{vv'}\setminus Q^{vv'}}(X_i-\lambda)\ .$$

Taking expectation,

$$\begin{aligned}&\mathbb{E}\left[\prod_{l=1}^{v}(X_{y_{\pi(l)}}-\lambda)\prod_{m=1}^{v'}(X_{y'_{\pi'(m)}}-\lambda)\right]\\ &= \mathbb{E}_{\pi_y,\pi_{y'}}\left[\mathbb{E}_{X_1,\ldots,X_s}\left[\prod_{l=1}^{v}(X_{y_{\pi(l)}}-\lambda)\prod_{m=1}^{v'}(X_{y'_{\pi'(m)}}-\lambda)\mid\pi_y,\pi_{y'}\right]\right]\\ &= \mathbb{E}_{\pi_y,\pi_{y'}}\left[\mathbb{E}_{X_1,\ldots,X_s}\left[\prod_{i\in Q^{vv'}}(X_i-\lambda)^2\prod_{i\in U^{vv'}\setminus Q^{vv'}}(X_i-\lambda)\Big|\pi_y,\pi_{y'}\right]\right]\\ &= \mathbb{E}_{\pi_y,\pi_{y'}}\left[\prod_{i\in Q^{vv'}}\mathbb{E}\big[(X_i-\lambda)^2\big]\prod_{i\in U^{vv'}\setminus Q^{vv'}}\mathbb{E}\big[X_i\big](X_i-\lambda)\Big|\pi_y,\pi_{y'}\right]\\ &= \mathbb{E}_{\pi_y,\pi_{y'}}\left[\eta^{2q^{vv'}}(\mu-\lambda)^{v+v'-2q^{vv'}}\right],\end{aligned}$$

by independence of the X_i's for $i\in[s]$.

Therefore,

$$\begin{aligned}&\mathsf{Cov}\left(\vartheta_y,\vartheta_{y'}\right)\\ &= \mathbb{E}\left[\vartheta_y\vartheta_{y'}\right]-\mathbb{E}\big[\vartheta_y\big]\mathbb{E}\big[\vartheta_{y'}\big]\\ &= \sum_{v,v'=0}^{k}\gamma_v(\lambda)\gamma_{v'}(\lambda)\left(\mathbb{E}\left[\prod_{l=1}^{v}(X_{y_{\pi(l)}}-\lambda)\prod_{m=1}^{v'}(X_{y'_{\pi'(m)}}-\lambda)\right]-(\mu-\lambda)^{v+v'}\right)\\ &= \sum_{v,v'=0}^{k}\gamma_v(\lambda)\gamma_{v'}(\lambda)\left(\mathbb{E}_{\pi_y,\pi_{y'}}\left[\eta^{2q^{vv'}}(\mu-\lambda)^{v+v'-2q^{vv'}}\right]-(\mu-\lambda)^{v+v'}\right)\\ &= \sum_{v,v'=1}^{k}\gamma_v(\lambda)\gamma_{v'}(\lambda)\left(\mathbb{E}_{\pi_y,\pi_{y'}}\left[\eta^{2q^{vv'}}(\mu-\lambda)^{v+v'-2q^{vv'}}\right]-(\mu-\lambda)^{v+v'}\right)\end{aligned}\tag{11}$$

where the last step follows by noting that if $v = 0$ or $v' = 0$, then $q^{vv'} = 0$ and so, $\eta^{2q^{vv'}}(\mu - \lambda)^{v+v'-2q^{vv'}} = (\mu-\lambda)^{v+v'}$. Hence the summation indices v, v' in (11) may start from 1 instead of 0.

Case 1: $\mu = \lambda$. If $v \neq v'$, then, $2q^{vv'} \leq 2\min(v, v') < v + v'$, Hence, the term $(\mu-\lambda)^{v+v'-2q^{vv'}} = 0$. In this case, Eqn. (11) becomes

$$\mathbb{E}\left[\vartheta_y \vartheta_{y'}\right] - \mathbb{E}\big[\vartheta_y\big]\mathbb{E}\big[\vartheta_{y'}\big] = \sum_{v=1}^{|y\cap y'|} \gamma_v^2(\lambda)\eta^{2v}\mathsf{Pr}_{\pi_y,\pi'_{y'}}\left[q^{vv}_{yy'\pi\pi'} = v\right] \tag{12}$$

Case 2: $\mu \neq \lambda$. Then, Eqn. (11) can be written as

$$\begin{aligned}&\mathbb{E}\left[\vartheta_y \vartheta_{y'}\right] - \mathbb{E}\big[\vartheta_y\big]\mathbb{E}\big[\vartheta_{y'}\big] \\ &= \sum_{v,v'=1}^{k} \gamma_v(\lambda)\gamma_{v'}(\lambda)(\mu-\lambda)^{v+v'}\left(\mathbb{E}_{\pi_y,\pi_{y'}}\left[\left(\frac{\eta^2}{(\mu-\lambda)^2}\right)^{q^{vv'}}\right] - 1\right) \quad .\end{aligned} \tag{13}$$

This proves the Lemma. □

Let Y be a code satisfying the properties of Corollary 5 and let $y, y' \in Y$ and distinct such that $t = |y \cap y'|$. Let $\pi_y, \pi_{y'}$ denote randomly and independently chosen permutations from $[k] \to [k]$. Define

$$P_{yy'} = \lambda^{2p}\sum_{v,v'=1}^{k}\binom{p}{v}\binom{p}{v'}\left(\frac{\mu-\lambda}{\lambda}\right)^{v+v'}\sum_{r=1}^{t}\left(\frac{\eta^2}{(\mu-\lambda)^2}\right)^{r}\mathsf{Pr}_{\pi_y,\pi_{y'}}\left[q^{vv'} = r\right] \tag{14}$$

$$Q_{yy'} = \lambda^{2p}\sum_{1\leq v,v'\leq k}\binom{p}{v}\binom{p}{v'}\left(\frac{\mu-\lambda}{\lambda}\right)^{v+v'}\left(\mathsf{Pr}_{\pi_y,\pi_{y'}}\left[q^{vv'} = 0\right] - 1\right) \tag{15}$$

Corollary 23. *Assume the premises and notation of Lemma 22 and let $\mu \neq \lambda$. For $y, y \in Y$ and $y \neq y'$ such that $t = |y \cap y'|$, let $\pi_y, \pi_{y'}$ denote randomly and independently chosen permutations from $[k] \to [k]$. Then,*

$$\mathsf{Cov}\left(\vartheta_y, \vartheta_{y'}\right) \leq P_{yy'} + Q_{yy'} \quad .$$

Proof. Since $\psi(x) = x^p$, $\gamma_v(\lambda) = \binom{p}{v}\lambda^{p-v}$. The Corollary follows by substituting this into Lemma 22. □

B.2 Probability of overlap of prefixes of y and y' after random ordering

Lemma 24. *Let Y be a code satisfying the properties of Corollary 5. Let $\{\pi_y\}_{y\in Y}$ be a family of random and independently chosen permutations from $[k] \to [k]$. For distinct $y, y' \in Y$,*

$$\mathsf{Pr}_{\pi_y,\pi_{y'}}\left[q^{vv'} = r\right] = \frac{1}{\binom{k}{v}\binom{k}{v'}}\sum_{s=0}^{t-r}\binom{t}{r}\binom{t-r}{s}\binom{k-t}{v-(r+s)}\binom{k-(r+s)}{v'-r} \tag{16}$$

Proof. Fix $y, y' \in Y$ and distinct and let $t = t(y, y') = |y \cap y'|$. By notation, $\pi_y(y)[v]$ is the v-sequence $\tau = (y_{\pi_y(1)}, \ldots, y_{\pi_y(v)})$ and $\pi_{y'}(y')[v]$ is the v'-sequence $\nu = (y'_{\pi'_y(1)}, \ldots, y'_{\pi'_y(v)})$. The permutations π_y and $\pi_{y'}$ are each uniformly randomly and independently chosen from the space of all permutations $[k] \to [k]$ (i.e., S_k).

The problem is to count the number of ways in which the v positions in τ and the v' positions in ν can be filled, using the elements of y and y' under permutations π_y and $\pi_{y'}$ such that $\tau \cap \nu$ has exactly r elements. Since π_y and $\pi_{y'}$ are uniformly random and independent permutations, the sample space has size $k^{\underline{v}} \cdot k^{\underline{v'}} = \binom{k}{v}\binom{k}{v'}v!v'!$. There are t elements in common among y and y' and we wish for τ and ν to have r elements in common. Suppose τ has $r + s$ elements from the t elements in common, where, s ranges from 0 to $\max(t - r, v - r)$. These are selected in $\binom{t}{r+s}$ ways. Having chosen these elements, we select r elements in $\binom{r+s}{r}$ ways–these elements are included in ν as well. We have now filled $r + s$ positions of τ and r positions of s. The remaining $v - (r + s)$ positions may be filled out of the $k - t$ elements of y that are not common with y'. This is done in $\binom{k-t}{v-(r+s)}$ ways. There are $v' - r$ positions remaining to be filled in ν. There are $k - t + (t - (r + s))$ elements to choose from, which can be done in $\binom{k-(r+s)}{v'-r}$ ways. The v elements chosen for τ and the v' elements chosen for ν can be rearranged in $v!$ and $v'!$ ways. Thus,

$$\begin{aligned}\mathsf{Pr}_{\pi_y,\pi_{y'}}\left[q^{vv'} = r\right] &= \frac{v!v'!}{\binom{k}{v}\binom{k}{v'}v!v'!}\sum_{s=0}^{t-r}\binom{t}{r+s}\binom{r+s}{r}\binom{k-t}{v-(r+s)}\binom{k-(r+s)}{v'-r} \\ &= \frac{1}{\binom{k}{v}\binom{k}{v'}}\sum_{s=0}^{t-r}\binom{t}{r}\binom{t-r}{s}\binom{k-t}{v-(r+s)}\binom{k-(r+s)}{v'-r} \qquad (17)\end{aligned}$$

which proves the lemma.

□

B.3 Estimating $Q_{yy'}$

Lemma 25. *Assume the premises and notation of Lemma 22 and Corollary 23. Let $p \geq 2$ and let $y, y' \in Y$ and distinct. If $\mu \neq \lambda$, then $Q_{yy'} < 0$.*

Proof. Fix $y, y' \in Y$ and distinct and let Q denote $Q_{yy'}$. Let $\alpha = \frac{\mu-\lambda}{\lambda} \leq \frac{1}{25p}$. Then,

$$Q = -Q_1 + Q_2$$

where,

$$Q_1 = \sum_{1\leq v,v'\leq k}\binom{p}{v}\binom{p}{v'}\lambda^{2p-v-v'}(\mu-\lambda)^{v+v'} = \lambda^{2p}\left(\sum_{v=1}^{k}\binom{p}{v}\alpha^v\right)^2 \qquad (18)$$

$$Q_2 = \sum_{1\leq v,v'\leq k}\binom{p}{v}\binom{p}{v'}\lambda^{2p-v-v'}(\mu-\lambda)^{v+v'}\mathsf{Pr}_{\pi,\pi'}\left[q^{vv'} = 0\right] \quad . \qquad (19)$$

Consider $\sum_{v=1}^{k}\binom{p}{v}\alpha^v$. The absolute value of the ratio of the $v+1$st term to the vth term, for $v = 1, 2, \ldots, k-1$, is

$$\frac{|p-v|}{v+1}\cdot\alpha \leq \left(\frac{p}{2}\right)\cdot\frac{1}{25p} \leq \frac{1}{50} \quad .$$

Therefore,

$$\left|\sum_{v=1}^{k}\binom{p}{v}\alpha^v - p\alpha\right| \le (p\alpha)\sum_{v\ge 1}(50)^{-(v-1)} = \frac{(p\alpha)}{49} \ .$$

Therefore,

$$Q_1 = \lambda^{2p}p\alpha\left(1\pm\frac{1}{49}\right)^2 \in \lambda^{2p}p\alpha\left(1\pm\frac{1}{24}\right) \tag{20}$$

Consider Q_2. Let $t = t(y,y') = |y\cap y'|$.

$$\begin{aligned} Q_2 &= \lambda^{2p}\sum_{v=1}^{k}\sum_{v'=1}^{k}\binom{p}{v}\binom{p}{v'}\alpha^{v+v'}\sum_{u=0}^{t}\frac{\binom{t}{u}\binom{k-t}{v-u}\binom{k-u}{v'}}{\binom{k}{v}\binom{k}{v'}} \\ &= \lambda^{2p}\sum_{u=0}^{t}\binom{t}{u}\sum_{v=1}^{k}\binom{p}{v}\frac{\binom{k-t}{v-u}}{\binom{k}{v}}\alpha^v\sum_{v'=1}^{k}\binom{p}{v'}\frac{\binom{k-u}{v'}}{\binom{k}{v'}}\alpha^{v'} \\ &= \lambda^{2p}\sum_{u=0}^{t}\binom{t}{u}R_{ut}S_{ut} \end{aligned} \tag{21}$$

where,

$$\begin{aligned} R_{ut} &= \sum_{v=1}^{k}\binom{p}{v}\frac{\binom{k-t}{v-u}}{\binom{k}{v}}\alpha^v = \sum_{v=\max(u,1)}^{k-t+u}\frac{\binom{p}{v}\binom{k-t}{v-u}}{\binom{k}{v}}\alpha^v, \quad \text{and} \\ S_{ut} &= \sum_{v'=1}^{k}\binom{p}{v'}\frac{\binom{k-u}{v'}}{\binom{k}{v'}}\alpha^{v'} \ . \end{aligned}$$

Consider R_{ut}. The absolute value of the ratio of the $(v+1)^{\text{st}}$ term in the summation R_{ut} to the vth term for $\max(u,1)\le v\le k-t+u-1$ is

$$\frac{|p-v|}{v+1}\cdot\left(\frac{k-t-v+u}{v-u+1}\right)\cdot\left(\frac{v+1}{k-v}\right)\alpha \le \left(\frac{p}{2}\right)\left(\frac{k-v-(t-u)}{k-v}\right)\cdot\frac{1}{25p} \le \frac{1}{50} \ .$$

Case 1: $u\le 1$. Then,

$$R_{ut} \in \frac{p\alpha}{k}\left(1\pm\frac{1}{49}\right) \ .$$

Case 2: $u\ge 2$. Then,

$$R_{ut} \in \frac{\binom{p}{u}\alpha^u}{\binom{k}{u}}\left(1\pm\frac{1}{49}\right) \ .$$

In either case,

$$R_{ut} \in \frac{\binom{p}{\max(u,1)}\alpha^{\max(u,1)}}{\binom{k}{\max(u,1)}}\left(1\pm\frac{1}{49}\right) \ . \tag{22}$$

Now consider $S_{ut} = \sum_{v=1}^{k} \binom{p}{v} \frac{\binom{k-u}{v}}{\binom{k}{v}} \alpha^v$. The absolute value of the ratio of the $v+1$th term in the summation S_{ut} to the vth term, $v = 1, 2, \ldots, k-u-1$ is

$$\frac{|p-v|}{v+1}\left(\frac{k-u-v}{v+1}\right)\left(\frac{v+1}{k-v}\right)\alpha \leq \frac{p\alpha}{2} \leq \frac{1}{50} \ .$$

Therefore,

$$S_{ut} \in \left(\frac{p(k-u)\alpha}{k}\right)\left(1 \pm \frac{1}{49}\right) \tag{23}$$

Substituting Eqns. (22) and (23) in Eqn. (21), we have,

$$\begin{aligned} Q_2 &= \lambda^{2p} \sum_{u=0}^{t} \binom{t}{u} R_{ut} S_{ut} \\ &\in \left(1 \pm \frac{1}{49}\right)\left(1 \pm \frac{1}{49}\right)\lambda^{2p} \sum_{u=0}^{t} \frac{\binom{t}{u}\binom{p}{\max(u,1)}\alpha^{\max(u,1)}(p\alpha)(k-u)}{\binom{k}{\max(u,1)}k} \end{aligned} \tag{24}$$

Consider the summation term in Eqn. (24).

$$\sum_{u=0}^{t} \frac{\binom{t}{u}\binom{p}{\max(u,1)}\alpha^{\max(u,1)}(k-u)}{\binom{k}{\max(u,1)}} = p\alpha + \sum_{u=1}^{t} \frac{\binom{t}{u}\binom{p}{u}\alpha^u(k-u)}{\binom{k}{u}} \tag{25}$$

Consider the summation term in Eqn. (25). The ratio of the absolute value of the $u+1$st term to the uth term, for $1 \leq u \leq t-1$ is

$$\left(\frac{t-u}{u+1}\right)\left(\frac{|p-u|}{u+1}\right)\left(\frac{u+1}{k-u}\right)\left(\frac{k-u-1}{k-u}\right)\alpha \leq \left(\frac{(t-u)}{k-u}\right)\left(\frac{p}{2}\right)(1)(\alpha) \leq \frac{1}{200}$$

since, $t \leq k/4$ from the property of the code Y.

Therefore, from Eqn. (25),

$$\sum_{u=1}^{t} \frac{\binom{t}{u}\binom{p}{u}\alpha^u(k-u)}{\binom{k}{u}} \in \frac{tp\alpha(k-1)}{k^2}\left(1 \pm \frac{1}{199}\right) \in \frac{p\alpha}{4}\left(1 \pm \frac{1}{199}\right)$$

since $t \leq k/4$.

Substituting in Eqn. (24), we have,

$$Q_2 \in \left(1 \pm \frac{1}{49}\right)^2 \left(\frac{p\alpha}{k}\right)\lambda^{2p}\left(p\alpha + \frac{p\alpha}{4}\left(1 \pm \frac{1}{199}\right)\right) \leq \left(\frac{(1.31)(p\alpha)^2}{k}\right)\lambda^{2p} \tag{26}$$

Using Eqns. (20) and (26), we have,

$$\begin{aligned} Q_1 - Q_2 &\geq \lambda^{2p}(p\alpha)^2\left(1 - \frac{1}{24}\right) - \lambda^{2p}\frac{(1.31)(p\alpha)^2}{k} \\ &> 0 \end{aligned}$$

since, $k \geq 3$. Hence, $Q = -Q_1 + Q_2 < 0$. □

B.4 Estimating $P_{yy'}$.

Notation. Let Y be a code satisfying Corollary 5. Let $y, y' \in Y$ and distinct and let $t = |y \cap y'|$. Let P denote $P_{yy'}$. Let $\alpha = \frac{\mu-\lambda}{\lambda}$ and $\beta = \frac{\eta^2}{\lambda^2}$. Define

$$P_1 = \lambda^{2p} \sum_{u=1}^{t} \binom{t}{u} \sum_{r=1}^{u} \binom{u}{r} \beta^r \left(\frac{\left|p^{\underline{u}}\right| \left|p^{\underline{r}}\right| |\alpha|^{u-r}}{k^{\underline{u}}\, k^{\underline{r}}} \right) \cdot \left((1-|\alpha|)^{p-u+p-r} + 2(1-|\alpha|)^{p-u}(27)^{-(3/4)k} + (27)^{-(1.5k)} \right) \mathbf{1}_{r>p, p \text{non-integral}}$$

$$P_2 = \lambda^{2p} \sum_{u=1}^{t} \binom{t}{u} \sum_{r=1}^{u} \binom{u}{r} \beta^r \left(\frac{\left|p^{\underline{u}}\right| \left|p^{\underline{r}}\right| |\alpha|^{u-r}}{k^{\underline{u}}\, k^{\underline{r}}} \right) \cdot \left((1-|\alpha|)^{p-u} + (27)^{-(3/4)k} \right) \left(\frac{50}{49} \right) \mathbf{1}_{r \le p < u, p \text{ non-integral}}$$

$$P_3 = \lambda^{2p} \sum_{u=1}^{t} \binom{t}{u} \sum_{r=1}^{u} \binom{u}{r} \beta^r \left(\frac{\left|p^{\underline{u}}\right| \left|p^{\underline{r}}\right| |\alpha|^{u-r}}{k^{\underline{u}}\, k^{\underline{r}}} \right) \cdot \left(\frac{50}{49} \right)^2 \mathbf{1}_{u \le p}$$

Lemma 26. *Assume the premises and notation of Lemma 22 and Corollary 23. Let $y, y' \in Y$ and distinct and let $\pi = \pi_y$ and $\pi' = \pi_{y'}$ be random permutations from $[k] \to [k]$. Let $\alpha = \frac{\mu-\lambda}{\lambda}$ and $\beta = \frac{\eta^2}{\lambda^2}$. Then, $P_{yy'} = 0$ if p is integral, and otherwise, $P_{yy'} \le P_1 + P_2 + P_3$.*

Proof. Let P denote $P_{yy'}$. Then,

$$\begin{aligned}
P &= \lambda^{2p} \sum_{v,v'=1}^{k} \binom{p}{v}\binom{p}{v'} \alpha^{v+v'-2r} \sum_{r=1}^{t} \beta^r \mathsf{Pr}_{\pi,\pi'}\left[q^{vv'} = r\right] \\
&= \lambda^{2p} \sum_{v,v'=1}^{k} \binom{p}{v}\binom{p}{v'} \sum_{r=1}^{t} \alpha^{v+v'-2r}\beta^r \cdot \frac{1}{\binom{k}{v}\binom{k}{v'}} \sum_{u=r}^{t} \binom{t}{u}\binom{u}{r}\binom{k-t}{v-u}\binom{k-u}{v'-r} \\
&= \lambda^{2p} \sum_{u=1}^{t} \binom{t}{u} \sum_{r=1}^{u} \binom{u}{r} \beta^r \left(\sum_{v=u}^{k} \binom{p}{v} \frac{\binom{k-t}{v-u}}{\binom{k}{v}} \cdot \alpha^{v-r} \right) \left(\sum_{v'=r}^{k} \binom{p}{v'} \frac{\binom{k-u}{v'-r}}{\binom{k}{v'}} \cdot \alpha^{v'-r} \right) \\
&= \lambda^{2p} \sum_{u=1}^{t} \binom{t}{u} \sum_{r=1}^{u} \binom{u}{r} \beta^r U_{ur} V_{ur} \qquad (27)
\end{aligned}$$

where,

$$U_{ur} = \sum_{v=u}^{k} \frac{\binom{p}{v}\binom{k-t}{v-u}\alpha^{v-r}}{\binom{k}{v}} \quad \text{and} \quad V_{ur} = \sum_{v'=r}^{k} \frac{\binom{p}{v'}\binom{k-u}{v'-r}\alpha^{v'-r}}{\binom{k}{v'}} \ . \qquad (28)$$

We first obtain upper bounds on U_{ur} and V_{ur}.

$$\begin{aligned}
U_{ur} &= \sum_{v=u}^{k} \frac{\binom{p}{v}\binom{k-t}{v-u}\alpha^{v-r}}{\binom{k}{v}} \\
&= \sum_{v=u}^{k} \frac{\frac{p^{\underline{u}}}{v^{\underline{u}}}\binom{p-u}{v-u}\binom{k-t}{v-u}\alpha^{v-u+(u-r)}}{\frac{k^{\underline{u}}}{v^{\underline{u}}}\binom{k-u}{v-u}} \\
&= \frac{p^{\underline{u}}\,\alpha^{u-r}}{k^{\underline{u}}} \sum_{w=0}^{k-u} \binom{p-u}{w} \frac{\binom{k-t}{w}}{\binom{k-u}{w}} \alpha^w \ . \qquad (29)
\end{aligned}$$

by letting $w = v - u$.

Case U.1: $u > p$. Note that if p is integral then $U_{ur} = 0$. Otherwise, $\text{sgn}(\binom{p-u}{w}) = (-1)^w$. Using this and since $0 \le t \le u$, we have,

$$\left|\sum_{w=0}^{k-u}\binom{p-u}{w}\alpha^w\frac{\binom{k-t}{w}}{\binom{k-u}{w}}\right| \le \sum_{w=0}^{k-u}\binom{p-u}{w}(-1)^w|\alpha|^w = (1-|\alpha|)^{p-u} + \binom{p-u}{k-u+1}\gamma^{k-u+1} \tag{30}$$

for some $\gamma \in (-|\alpha|, 0)$, by Taylor's series expansion of $(1-|\alpha|)^{p-u}$ around 0 up to $k-u$ terms.

Now, for $u > p$, $1 \le u \le t \le k/4$, we have,

$$\left|\binom{p-u}{k-u+1}\gamma^{k-u+1}\right| \le \binom{k-p}{k-u+1}|\alpha|^{k-u+1} \le \left(\frac{(k-p)e|\alpha|}{k-u+1}\right)^{k-u+1} \le (27)^{-(3/4)k} \ . \tag{31}$$

since, $1 \le u \le t \le k/4$ and $|\alpha| \le \frac{1}{25p} \le \frac{1}{50}$.

Case U.2: $u \le p$. Consider $\sum_{w=0}^{k-u}\binom{p-u}{w}\alpha^w\frac{\binom{k-t}{w}}{\binom{k-u}{w}}$. Let the wth term in the summation be τ_w, for $0 \le w \le k-u-1$. Then, for $1 \le w \le k-u-1$,

$$\left|\frac{\tau_{w+1}}{\tau_w}\right| = \left(\frac{|p-u-w|}{w+1}\right)\cdot|\alpha|\cdot\left(\frac{k-t-w}{k-u-w}\right) \le \frac{1}{50}$$

since, (a) $1 \le u \le t$ and $k-u-w \ge 1$, and, (b) $\frac{|(p-u)-w|}{w+1} \le \frac{p}{2}$.

Therefore,

$$\left|\sum_{w=0}^{k-u}\binom{p-u}{w}\alpha^w\frac{\binom{k-t}{w}}{\binom{k-u}{w}} - 1\right| \le \sum_{w\ge 1}(50)^{-w} = \frac{1}{49} \ .$$

Combining Cases U.1 and U.2, we have,

$$|U_{ur}| \le \left(\frac{|p^{\underline{u}}|\,|\alpha|^{u-r}}{k^{\underline{u}}}\right)\left[\left((1-|\alpha|)^{p-u} + (27)^{-(3/4)k}\right)\mathbf{1}_{u>p,p \text{ non-integral}} + \frac{50}{49}\mathbf{1}_{u\le p}\right] \ . \tag{32}$$

Case V: Proceeding similarly for evaluating V_{ur}, we have,

$$\begin{aligned} V_{ur} &= \sum_{v=r}^{k}\frac{\binom{p}{v}\binom{k-u}{v-r}\alpha^{v-r}}{\binom{k}{v}} \\ &= \sum_{v=r}^{k}\frac{\frac{p^{\underline{r}}}{v^{\underline{r}}}\binom{p-r}{v-r}\binom{k-u}{v-r}\alpha^{v-r}}{\frac{k^{\underline{r}}}{v^{\underline{r}}}\binom{k-r}{v-r}} \\ &= \frac{p^{\underline{r}}}{k^{\underline{r}}}\sum_{w=0}^{k-r}\frac{\binom{p-r}{w}\binom{k-u}{w}\alpha^w}{\binom{k-r}{w}} \ . \end{aligned}$$

Case V.1: $r > p$. We note that if p is integral then $p^{\underline{r}} = 0$ and therefore $V_{ur} = 0$. Otherwise, $\text{sgn}(\binom{p-r}{w}) = (-1)^w$. Thus,

$$\begin{aligned}\left|\sum_{w=0}^{k-r}\frac{\binom{p-r}{w}\binom{k-u}{w}\alpha^w}{\binom{k-r}{w}}\right| &\le \sum_{w=0}^{k-r}\frac{\left|\binom{p-r}{w}\right||\alpha|^w\binom{k-u}{w}}{\binom{k-r}{w}}\\ &\le \sum_{w=0}^{k-r}\left|\binom{p-r}{w}\right||\alpha|^w, \text{ since, } k\ge u\ge r\ge 1,\\ &= \sum_{w=0}^{k-r}\binom{p-r}{w}(-|\alpha|)^w, \text{for some } \gamma\in(-|\alpha|,0),\\ &= (1-|\alpha|)^{p-r}+\binom{p-r}{k-r+1}\gamma^{k-r+1}\\ &\le (1-|\alpha|)^{p-r}+(27)^{-(3/4)k}\end{aligned}$$

following the same argument as in Eqn. (31), and using $1\le r\le t\le k/4$. Thus,

$$\left|V_{ur}\right| \le \frac{|p^{\underline{r}}|}{k^{\underline{r}}}\left((1-|\alpha|)^{p-r}+(27)^{-(3/4)k}\right)$$

Case V.2: $r\le p$. Consider the ratio of the absolute value of the $w+1$st term, denoted ν_{w+1} to the wth term ν_w of the summation $\sum_{w=0}^{k-r}\binom{p-r}{w}\alpha^w\frac{\binom{k-u}{w}}{\binom{k-r}{w}}$. Then,

$$\left|\frac{\nu_{w+1}}{\nu_w}\right| = \left(\frac{|p-r-w|}{w+1}\right)\alpha\left(\frac{k-u-w}{k-r-w}\right) \le \left(\frac{p}{2}\right)\alpha \le \frac{1}{50}\ .$$

Therefore,

$$\sum_{w=0}^{k-r}\binom{p-r}{w}\alpha^w\frac{\binom{k-u}{w}}{\binom{k-r}{w}} \in \left(1\pm\frac{1}{49}\right)\ .$$

and so,

$$|V_{ur}| \in \frac{p^{\underline{r}}}{k^{\underline{r}}}\left(1\pm\frac{1}{49}\right)\ .$$

Combining Cases *V.1* and *V.2* gives

$$|V_{ur}| \le \frac{|p^{\underline{r}}|}{k^{\underline{r}}}\left(\left((1-|\alpha|)^{p-r}+(27)^{-(3/4)k}\right)\mathbf{1}_{r>p,p\text{ non-integral}}+\frac{50}{49}\cdot\mathbf{1}_{r\le p}\right) \tag{33}$$

Substituting Eqn. (32) and (33) in Eqn. (51), we have,

$$\begin{aligned}P &= \lambda^{2p}\sum_{u=1}^{t}\binom{t}{u}\sum_{r=1}^{u}\binom{u}{r}\beta^r U_{u,r}V_{u,r}\\ &\le \lambda^{2p}\sum_{u=1}^{t}\binom{t}{u}\sum_{r=1}^{u}\binom{u}{r}\beta^r\left|U_{u,r}\right|\left|V_{u,r}\right| \end{aligned}\tag{34}$$

Now, since, $1 \le r \le u \le t \le k/4$, we have,

$$
\begin{aligned}
&\left|U_{ur}\right| \cdot \left|V_{ur}\right| \\
&\le \frac{\left|p^{\underline{u}}\right| |\alpha|^{u-r}}{k^{\underline{u}}} \left(\left((1-|\alpha|)^{p-u} + (27)^{-(3/4)k} \right) \mathbf{1}_{u>p,p \text{ non-integral}} + \frac{50}{49} \cdot \mathbf{1}_{u \le p} \right) \\
&\quad \cdot \left(\frac{\left|p^{\underline{r}}\right|}{k^{\underline{r}}} \right) \left(\left((1-|\alpha|)^{p-r} + (27)^{-(3/4)k} \right) \mathbf{1}_{r>p,p \text{ non-integral}} + \frac{50}{49} \cdot \mathbf{1}_{r \le p} \right) \\
&\le \left(\frac{\left|p^{\underline{u}}\right| \left|p^{\underline{r}}\right| |\alpha|^{u-r}}{k^{\underline{u}}\, k^{\underline{r}}} \right) \left(\left((1-|\alpha|)^{p-u+p-r} + 2(1-|\alpha|)^{p-u}(27)^{-(3/4)k} + (27)^{-(1.5k)} \right) \mathbf{1}_{r>p,p \text{ non-integral}} \right. \\
&\qquad \left. + \left((1-|\alpha|)^{p-u} + (27)^{-(3/4)k} \right) \left(\frac{50}{49} \right) \mathbf{1}_{r \le p < u,p \text{ non-integral}} \right) + \left(\frac{50}{49} \right)^2 \mathbf{1}_{u \le p} \Bigg)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
P &= \lambda^{2p} \sum_{u=1}^{t} \binom{t}{u} \sum_{r=1}^{u} \binom{u}{r} \beta^r U_{ur} \cdot V_{ur} \\
&\le \lambda^{2p} \sum_{u=1}^{t} \binom{t}{u} \sum_{r=1}^{u} \binom{u}{r} \beta^r \left(\frac{\left|p^{\underline{u}}\right| \left|p^{\underline{r}}\right| |\alpha|^{u-r}}{k^{\underline{u}}\, k^{\underline{r}}} \right) \\
&\qquad \cdot \left(\left((1-|\alpha|)^{p-u+p-r} + 2(1-|\alpha|)^{p-u}(27)^{-(3/4)k} + (27)^{-(1.5k)} \right) \mathbf{1}_{r>p,p \text{ non-integral}} \right. \\
&\qquad\qquad \left. + \left((1-|\alpha|)^{p-u} + (27)^{-(3/4)k} \right) \left(\frac{50}{49} \right) \mathbf{1}_{r \le p < u,p \text{ non-integral}} \right) + \left(\frac{50}{49} \right)^2 \mathbf{1}_{u \le p} \Bigg) \\
&= P_1 + P_2 + P_3
\end{aligned}
$$

□

B.4.1 Estimating P_3

Lemma 27. *Assume the premises and notation of Lemma 22 and Corollary 23. Let $y, y' \in Y$ and distinct and let $\pi = \pi_y$ and $\pi' = \pi_{y'}$ be random permutations from $[k] \to [k]$. Let $\alpha = \frac{\mu - \lambda}{\lambda}$ and $\beta = \frac{\eta^2}{\lambda^2}$. Then,*

$$
P_3 \le \frac{0.275p^2}{k} \lambda^{2p} \beta \ . \tag{35}
$$

Proof. Consider the sum P_3.

$$\begin{aligned}
P_3 &= \lambda^{2p} \sum_{u=1}^{t} \binom{t}{u} \sum_{r=1}^{u} \binom{u}{r} \beta^r \left(\frac{|p^{\underline{u}}|\,|p^{\underline{r}}|\,|\alpha|^{u-r}}{k^{\underline{u}}\, k^{\underline{r}}} \right) \cdot \left(\frac{50}{49}\right)^2 \mathbf{1}_{u \le p} \\
&= \left(\frac{50}{49}\right)^2 \lambda^{2p} \sum_{u=1}^{\min(p,t)} \binom{t}{u} \left(\frac{p^{\underline{u}}}{k^{\underline{u}}}\right) |\alpha|^u \sum_{r=1}^{u} \binom{u}{r} \left(\frac{p^{\underline{r}}}{k^{\underline{r}}}\right) \left(\frac{\beta}{|\alpha|}\right)^r \\
&\le \left(\frac{50}{49}\right)^2 \lambda^{2p} \sum_{u=1}^{\min(p,t)} \binom{t}{u} \left(\frac{p}{k}\right)^u |\alpha|^u \sum_{r=1}^{u} \binom{u}{r} \left(\frac{p}{k}\right)^r \left(\frac{\beta}{|\alpha|}\right)^r \\
&= \left(\frac{50}{49}\right)^2 \lambda^{2p} \sum_{u=1}^{\min(p,t)} \binom{t}{u} \left(\frac{p}{k}\right)^u |\alpha|^u \left(\left(1 + \frac{p\beta}{k|\alpha|}\right)^u - 1 \right) \\
&\le \left(\frac{50}{49}\right)^2 \lambda^{2p} \sum_{u=1}^{t} \binom{t}{u} \left(\frac{p}{k}\right)^u |\alpha|^u \left(\left(1 + \frac{p\beta}{k|\alpha|}\right)^u - 1 \right) \\
&= \left(\frac{50}{49}\right)^2 \lambda^{2p} \left(P_{31} - P_{32}\right) \qquad (36)
\end{aligned}$$

where,

$$\begin{aligned}
P_{31} &= \sum_{u=1}^{t} \binom{t}{u} \left(\frac{p}{k}\right)^u |\alpha|^u \left(1 + \frac{p\beta}{k|\alpha|}\right)^u = \left(1 + \frac{p|\alpha|}{k}\left(1 + \frac{p\beta}{k|\alpha|}\right)\right)^t - 1 \\
P_{32} &= \sum_{u=1}^{t} \binom{t}{u} |\alpha|^u \left(\frac{p}{k}\right)^u = \left(1 + \frac{p|\alpha|}{k}\right)^t - 1 \ .
\end{aligned}$$

Let $a = \left(1 + \frac{p|\alpha|}{k}\left(1 + \frac{p\beta}{k|\alpha|}\right)\right) \le \exp\left\{\frac{p|\alpha|}{k}\left(1 + \frac{p\beta}{k|\alpha|}\right)\right\}$ and $b = \left(1 + \frac{p|\alpha|}{k}\right)$. Therefore,

$$\begin{aligned}
P_{31} - P_{32} = a^t - b^t &\le (a-b)(ta^{t-1}) \\
&\le \left(\frac{p^2\beta}{k^2}\right)(t)\, \exp\left\{(t-1)\frac{p|\alpha|}{k}\left(1 + \frac{p\beta}{k|\alpha|}\right)\right\} \\
&\le \frac{p^2\beta}{4k}\, \exp\left\{\frac{p|\alpha|}{4} + \frac{p^2\beta}{4k}\right\} \\
&\le \frac{p^2\beta}{4k}\, \exp\left\{\frac{1}{100} + \frac{1}{50k}\right\} \\
&\le \frac{(1.0102)p^2\beta}{4k}
\end{aligned}$$

Therefore, subsituting in Eqn. (36), we have,

$$P_3 \le \frac{0.275p^2}{k}\lambda^{2p}\beta = \frac{0.275p^2}{k}\lambda^{2p-2}\eta^2 \ .$$

□

B.4.2 Estimating P_2

We now consider P_2.

Lemma 28. *Assume the premises and notation of Lemma 22 and Corollary 23. Let $y, y' \in Y$ and distinct and let $\pi = \pi_y$ and $\pi' = \pi_{y'}$ be random permutations from $[k] \to [k]$. Let $\alpha = \frac{\mu-\lambda}{\lambda}$ and $\beta = \frac{\eta^2}{\lambda^2}$.*

$$P_2 \leq \frac{p^2 \lambda^{2p} \beta}{(30)(40)k} \ . \tag{37}$$

Proof.

$$\begin{aligned} P_2 &= \lambda^{2p} \sum_{u=1}^{t} \binom{t}{u} \sum_{r=1}^{u} \binom{u}{r} \beta^r \left(\frac{\lfloor p^{\underline{u}} \rfloor \lfloor p^{\underline{r}} \rfloor |\alpha|^{u-r}}{k^{\underline{u}}\, k^{\underline{r}}} \right) \cdot \left((1-|\alpha|)^{p-u} + (27)^{-(3/4)k} \right) \left(\frac{50}{49} \right) \mathbf{1}_{r \leq p < u} \\ &= \left(\frac{50}{49} \right) \lambda^{2p} \sum_{u=\lfloor p \rfloor + 1}^{t} \binom{t}{u} \sum_{r=1}^{\lfloor p \rfloor} \binom{u}{r} \beta^r \left(\frac{\lfloor p^{\underline{u}} \rfloor \lfloor p^{\underline{r}} \rfloor |\alpha|^{u-r}}{k^{\underline{u}}\, k^{\underline{r}}} \right) \cdot \left((1-|\alpha|)^{p-u} + (27)^{-(3/4)k} \right) \end{aligned} \tag{38}$$

The first summation is empty if $t < \lfloor p \rfloor + 1$ in which case $P_2 = 0$. Also, $P_2 = 0$ if p is integral, since $p^{\underline{u}} = 0$, for $u \geq \lfloor p \rfloor + 1$. So we now assume that $t \geq \lfloor p \rfloor + 1$ and p is not integral. Further, $(1-|\alpha|)^{p-u} \geq 1$, for $u \geq \lfloor p \rfloor + 1$ and $|\alpha| \leq 1/(50p)$. Hence, $(27)^{-(3/4)k} + (1-|\alpha|)^{p-u} \leq (1-|\alpha|)^{p-u}(1+(27)^{-(3/4)k}$. Using this simplification and also using the fact that $p^{\underline{r}}/k^{\underline{r}} \leq (p/k)^r$, for $1 \leq r \leq \lfloor p \rfloor$, Eqn. (38) can be written as follows.

$$\begin{aligned} P_2 &\leq \left(\frac{50}{49} \right) \left(1 + (27)^{-(3/4)k} \right) \lambda^{2p} \sum_{u=\lfloor p \rfloor + 1}^{t} \binom{t}{u} \frac{\lfloor p^{\underline{u}} \rfloor |\alpha|^u}{k^{\underline{u}}} \sum_{r=1}^{\lfloor p \rfloor} \binom{u}{r} \left(\frac{\beta}{|\alpha|} \right)^r \left(\frac{\lfloor p^{\underline{r}} \rfloor}{k^{\underline{r}}} \right) \cdot (1-|\alpha|)^{p-u} \\ &\leq (1.042)(1-|\alpha|)^p \lambda^{2p} \sum_{u=\lfloor p \rfloor + 1}^{t} \sum_{r=1}^{\lfloor p \rfloor} \binom{t}{u} \binom{u}{r} \frac{\lfloor p^{\underline{u}} \rfloor \gamma^u}{k^{\underline{u}}} \left(\frac{\beta}{|\alpha|} \right)^r \left(\frac{p}{k} \right)^r \end{aligned}$$

where, $\gamma = \frac{|\alpha|}{1-|\alpha|}$.

Let

$$Q_2 = \sum_{u=\lfloor p \rfloor + 1}^{t} \sum_{r=1}^{\lfloor p \rfloor} \binom{t}{u} \binom{u}{r} \frac{\lfloor p^{\underline{u}} \rfloor \gamma^u}{k^{\underline{u}}} \left(\frac{\beta}{|\alpha|} \right)^r \left(\frac{p}{k} \right)^r$$

so that

$$P_2 \leq (1.042) e^{-|\alpha|p} \lambda^{2p} Q_2 \leq (1.001) \lambda^{2p} Q_2 \ . \tag{39}$$

Then,

$$
\begin{aligned}
Q_2 &= \sum_{u=\lfloor p\rfloor+1}^{t}\sum_{r=1}^{\lfloor p\rfloor}\binom{t}{u}\binom{u}{r}\frac{\left|p^{\underline{u}}\right|\gamma^u}{k^{\underline{u}}}\left(\frac{\beta}{|\alpha|}\right)^r\left(\frac{p}{k}\right)^r \\
&= \sum_{u=\lfloor p\rfloor+1}^{t}\sum_{r=1}^{\lfloor p\rfloor}\binom{t}{r}\binom{t-r}{u-r}\frac{\left|p^{\underline{u}}\right|\gamma^u}{k^{\underline{u}}}\left(\frac{\beta}{|\alpha|}\right)^r\left(\frac{p}{k}\right)^r \\
&= \sum_{r=1}^{\lfloor p\rfloor}\binom{t}{r}\left(\frac{p\beta}{|\alpha|k}\right)^r\sum_{u=\lfloor p\rfloor+1}^{t}\binom{t-r}{u-r}\frac{p^{\underline{r}}\left|(p-r)^{\underline{u-r}}\right|}{k^{\underline{r}}\,(k-r)^{\underline{u-r}}}\gamma^{r+u-r} \\
&= \sum_{r=1}^{\lfloor p\rfloor}\binom{t}{r}\left(\frac{p\beta}{|\alpha|k}\right)^r\left(\frac{p^{\underline{r}}}{k^{\underline{r}}}\right)\gamma^r\sum_{u-r=\lfloor p\rfloor+1-r}^{t-r}\binom{t-r}{u-r}\frac{\left|(p-r)^{\underline{u-r}}\right|}{(k-r)^{\underline{u-r}}}\cdot\gamma^{u-r} \qquad (40)
\end{aligned}
$$

Consider the inner summation in Eqn. (40), namely,

$$
\sum_{w=\lfloor p\rfloor+1-r}^{t-r}\binom{t-r}{w}\frac{\left|(p-r)^{\underline{w}}\right|}{(k-r)^{\underline{w}}}\cdot\gamma^w \qquad (41)
$$

The ratio of $(w+1)$st term to the wth term, for $w=\lfloor p\rfloor+1-r,\ldots,t-r-1$, in the above summation is

$$
\left(\frac{t-r-w}{w+1}\right)\left(\frac{|p-r-w|}{k-r-w}\right)\gamma=\left(\frac{t-r-w}{k-r-w}\right)\left(\frac{w-(p-r)}{w+1}\right)\gamma\le\frac{t\gamma}{k}\le\frac{1}{(4)(50p-1)}\le\frac{1}{(4)(49)}
$$

since $t\le k/4$ and $\gamma=\frac{|\alpha|}{1-|\alpha|}\le\frac{1}{50p-1}\le\frac{1}{49}$. Therefore, Eqn. (41) may be upper bounded as follows.

$$
\begin{aligned}
&\sum_{w=\lfloor p\rfloor+1-r}^{t-r}\binom{t-r}{w}\frac{\left|(p-r)^{\underline{w}}\right|}{(k-r)^{\underline{w}}}\cdot\gamma^w \\
&\le\binom{t-r}{\lfloor p\rfloor+1-r}\left(\frac{\left|(p-r)^{\underline{\lfloor p\rfloor+1-r}}\right|}{(k-r)^{\underline{\lfloor p\rfloor+1-r}}}\right)\gamma^{\lfloor p\rfloor+1-r}\left(1+\frac{1}{195}\right)\ .
\end{aligned}
$$

Substituting in Eqn. (40), we have,

$$
\begin{aligned}
Q_2 &\le (1.0052)\sum_{r=1}^{\lfloor p\rfloor}\binom{t}{r}\left(\frac{p\beta}{|\alpha|k}\right)^r\left(\frac{p^{\underline{r}}}{k^{\underline{r}}}\right)\gamma^r\binom{t-r}{\lfloor p\rfloor+1-r}\left(\frac{\left|(p-r)^{\underline{\lfloor p\rfloor+1-r}}\right|}{(k-r)^{\underline{\lfloor p\rfloor+1-r}}}\right)\gamma^{\lfloor p\rfloor+1-r} \\
&\le (1.0052)\sum_{r=1}^{\lfloor p\rfloor}\binom{t}{r}\binom{t-r}{\lfloor p\rfloor+1-r}\left(\frac{p^2\beta\gamma}{|\alpha|k^2}\right)^r\left(\frac{\left|(p-r)^{\underline{\lfloor p\rfloor+1-r}}\right|}{(k-r)^{\underline{\lfloor p\rfloor+1-r}}}\right)\gamma^{\lfloor p\rfloor+1-r} \\
&= (1.0052)\sum_{r=1}^{\lfloor p\rfloor}\binom{t}{\lfloor p\rfloor+1}\binom{\lfloor p\rfloor+1}{r}\left(\frac{p^2\beta}{(1-|\alpha|)k^2}\right)^r\left(\frac{\left|(p-r)^{\underline{\lfloor p\rfloor+1-r}}\right|}{(k-r)^{\underline{\lfloor p\rfloor+1-r}}}\right)\gamma^{\lfloor p\rfloor+1-r} \\
&= (1.0052)S \qquad (42)
\end{aligned}
$$

Consider the summation above and let $t_r = \binom{\lfloor p \rfloor + 1}{r} \left(\frac{p^2\beta}{(1-|\alpha|)k^2}\right)^r \left(\frac{(p-r)^{\underline{\lfloor p \rfloor + 1 - r}}}{(k-r)^{\underline{\lfloor p \rfloor + 1 - r}}}\right) \gamma^{\lfloor p \rfloor + 1 - r}$ be the rth term. Let $r_m = \text{argmax}_{r=1}^{\lfloor p \rfloor} t_r$, that is t_{r_m} is the largest among the t_r's. Then, clearly, $S = \sum_{r=1}^{\lfloor p \rfloor} t_r \leq \lfloor p \rfloor t_{r_m}$. For $r_m = r \in \{1, 2, \ldots, \lfloor p \rfloor\}$, we have,

$$\begin{aligned} S &\leq \lfloor p \rfloor \binom{t}{\lfloor p \rfloor + 1} \binom{\lfloor p \rfloor + 1}{r} \left(\frac{p^2\beta}{(1-|\alpha|)k^2}\right)^r \left(\frac{(p-r)^{\underline{\lfloor p \rfloor + 1 - r}}}{(k-r)^{\underline{\lfloor p \rfloor + 1 - r}}}\right) \gamma^{\lfloor p \rfloor + 1 - r} \\ &= \left(\frac{\lfloor p \rfloor \gamma}{r!}\right) \left(\frac{t^{\underline{\lfloor p \rfloor + 1}}}{k^r (k-r)^{\underline{\lfloor p \rfloor + 1 - r}}}\right) \left(\frac{p^2\beta}{(1-|\alpha|)k}\right)^r \left(\frac{(p-r)^{\underline{\lfloor p \rfloor + 1 - r}}}{(\lfloor p \rfloor + 1 - r)!}\right) \gamma^{\lfloor p \rfloor - r} \end{aligned} \tag{43}$$

Now

$$\frac{p^2\beta}{(1-|\alpha|)k} \leq \frac{p^2\beta}{(1-\frac{1}{50p})k} \leq \frac{(1.011)p^2\beta}{k}$$

since $p \geq 2$.

Therefore, Eqn. (43) may be written as

$$\begin{aligned} S &\leq \left(\frac{\lfloor p \rfloor \gamma}{r!}\right) \left(\frac{t}{k}\right)^r \left(\frac{t-r}{k-r}\right)^{\lfloor p \rfloor + 1 - r} \left(\frac{(1.011)p^2\beta}{k}\right)^r \gamma^{\lfloor p \rfloor - r} \\ &\leq \left(\frac{(1.011)\lfloor p \rfloor \gamma p^2 \beta}{k^r r!}\right) (4)^{-(\lfloor p \rfloor + 1)}, \text{ since, } \frac{t}{k} \leq \frac{1}{4} \text{ and } \gamma \leq \frac{1}{(50p-1)} \\ &\leq \frac{p^2\beta}{(30)(49)k}, \text{ since, } p \geq 2 \text{ and } p^2\beta \ll 1. \end{aligned} \tag{44}$$

Substituting in Eqn. (42), we have that $Q_2 \leq (1.0052)S$ and from Eqn. (39), we have,

$$P_2 \leq (1.001)\lambda^{2p} Q_2 \leq \frac{p^2\lambda^{2p}\beta}{(30)(40)k}$$

□

B.4.3 Estimating P_1

We now calculate P_1.

Lemma 29. *Assume the premises and notation of Lemma 22 and Corollary 23. Let $y, y' \in Y$ and distinct and let $\pi = \pi_y$ and $\pi' = \pi_{y'}$ be random permutations from $[k] \to [k]$. Let $\alpha = \frac{\mu - \lambda}{\lambda}$ and $\beta = \frac{\eta^2}{\lambda^2}$. Then for $n \geq 2$,*

$$P_1 \leq (0.3) \left(\frac{p^2\beta}{k^a}\right) \left(\frac{1}{(2)(25)^2 p}\right)^{(a-1)} \tag{45}$$

Proof.

$$\begin{aligned} P_1 = \lambda^{2p} \sum_{u=1}^{t} \binom{t}{u} \sum_{r=1}^{u} \binom{u}{r} \beta^r \left(\frac{\left|p^{\underline{u}}\right| \left|p^{\underline{r}}\right| |\alpha|^{u-r}}{k^{\underline{u}}\, k^{\underline{r}}}\right) \\ \cdot \left((1-|\alpha|)^{p-u+p-r} + 2(1-|\alpha|)^{p-u}(27)^{-(3/4)k} + (27)^{-(1.5k)}\right) \mathbf{1}_{r>p} \end{aligned} \tag{46}$$

First, we note that for $k \geq c \log n$ (where, $c = 100$ as per Table 2), $(27)^{(-3/4)k} = n^{-(3.5)c} \leq n^{-(3.5)c}(1-|\alpha|)^{p-u}$. Hence, $\left((1-|\alpha|)^{p-u+p-r} + 2(1-|\alpha|)^{p-u}(27)^{-(3/4)k} + (27)^{-(1.5k)}\right) = (1-|\alpha|)^{2p-u-r}(1+O(n^{-(3.5c)})$. Therefore,

$$\begin{aligned}
P_1 &= \left(1+O(n^{-3.5c})\right)\lambda^{2p}\sum_{u=1}^{t}\binom{t}{u}\sum_{r=1}^{u}\binom{u}{r}\beta^r\left(\frac{\lfloor p^{\underline{u}}\rfloor\lfloor p^{\underline{r}}\rfloor|\alpha|^{u-r}}{k^{\underline{u}}\,k^{\underline{r}}}\right)(1-|\alpha|)^{p-u+p-r}\mathbf{1}_{r>p} \\
&= \left(1+O(n^{-3.5c})\right)\lambda^{2p}\sum_{u=\lfloor p\rfloor+1}^{t}\sum_{r=\lfloor p\rfloor+1}^{u}\binom{t}{u}\binom{u}{r}\beta^r\left(\frac{\lfloor p^{\underline{u}}\rfloor\lfloor p^{\underline{r}}\rfloor|\alpha|^{u-r}}{k^{\underline{u}}\,k^{\underline{r}}}\right)(1-|\alpha|)^{2p-u-r} \\
&= \left(1+O(n^{-3.5c})\right)\lambda^{2p}L \qquad (47)
\end{aligned}$$

where,

$$L = \sum_{u=\lfloor p\rfloor+1}^{t}\sum_{r=\lfloor p\rfloor+1}^{u}\binom{t}{u}\binom{u}{r}\beta^r\left(\frac{\lfloor p^{\underline{u}}\rfloor\lfloor p^{\underline{r}}\rfloor|\alpha|^{u-r}}{k^{\underline{u}}\,k^{\underline{r}}}\right)(1-|\alpha|)^{2p-u-r} \ .$$

Let $a = \lfloor p\rfloor + 1$ and let $v = u - a$ and $w = r - a$. Then,

$$\begin{aligned}
L = (1-|\alpha|)^{2p-2a}\beta^a t^{\underline{a}}\left(\frac{p^{\underline{a}}}{k^{\underline{a}}}\right)^2\sum_{v=0}^{t-a}\binom{t-a}{v}\sum_{w=0}^{v}\binom{v}{w}|\alpha|^{v-w} \\
\left(\frac{\beta}{(1-|\alpha|)}\right)^w\left(\frac{(a+v-p-1)^{\underline{v}}\,(a+w-p-1)^{\underline{w}}}{(k-a)^{\underline{v}}\,(k-a)^{\underline{w}}\,(w+a)^{\underline{a}}}\right)(1-|\alpha|)^{-v} \qquad (48)
\end{aligned}$$

Now $a-p+w-1^{\underline{w}} \leq w^{\underline{w}} = w!$. Similarly, $a-p+v-1^{\underline{\leq}}\, v!$. Therefore,

$$\binom{t-a}{v}\binom{v}{w}(a+v-p-1)^{\underline{v}}\,(a+w-p-1)^{\underline{w}} \leq (t-a)^{\underline{v}}\,v^{\underline{w}} \ .$$

Hence,

$$\begin{aligned}
L &\leq (1-|\alpha|)^{2p-2a}\beta^a t^{\underline{a}}\left(\frac{p^{\underline{a}}}{k^{\underline{a}}}\right)^2\sum_{v=0}^{t-a}\left(\frac{(t-a)^{\underline{v}}}{(k-a)^{\underline{v}}}\right)(1-|\alpha|)^{-v} \\
&\qquad \sum_{w=0}^{v}\frac{v^{\underline{w}}}{(k-a)^{\underline{w}}}|\alpha|^{v-w}\left(\frac{\beta}{(1-|\alpha|)}\right)^w\left(\frac{1}{(w+a)^{\underline{a}}}\right) \\
&\leq (1-|\alpha|)^{2p-2a}\beta^a t^{\underline{a}}\left(\frac{p^{\underline{a}}}{k^{\underline{a}}}\right)^2\left(1+\sum_{v=1}^{t-a}\sum_{w=0}^{v}\frac{c^v v^{\underline{w}}\,|\alpha|^{v-w}}{(k-a)^{\underline{w}}}\beta'^w\left(\frac{1}{(w+a)^{\underline{a}}}\right)\right) \qquad (49)
\end{aligned}$$

where $c = \frac{t-a}{(k-a)(1-|\alpha|)}$ and $\beta' = \left(\frac{\beta}{(1-|\alpha|)}\right)$. Let l_{vw} denote the summand

$$l_{vw} = \frac{c^v v^{\underline{w}}\,|\alpha|^{v-w}}{(k-a)^{\underline{w}}}\beta'^w\left(\frac{1}{(w+a)^{\underline{a}}}\right), \quad 1 \leq v \leq t-a, 0 \leq w \leq v \ .$$

The summation in Eqn. (49) may be written as

$$J = \sum_{v=1}^{t-a}K_v, \text{ where, } K_v = \sum_{w=0}^{v}l_{vw}, \quad v = 1,2,\ldots,t-a \ .$$

Therefore,

Comparing l_{vw} and $l_{v+1,w}$, we have,

$$l_{vw} = \frac{c^v v^{\underline{w}} \, |\alpha|^{v-w} \beta'^w}{(k-a)^{\underline{w}} \, (w+a)^{\underline{a}}}$$
$$l_{v+1,w} = \frac{c^{v+1} (v+1)^{\underline{w}} \, |\alpha|^{v+1-w} \beta'^w}{(k-a)^{\underline{w}} \, (w+a)^{\underline{a}}}$$

Then,

$$\frac{l_{v+1,w+1}}{l_{vw}} = \frac{c(v+1)|\alpha|\beta'(w+1)}{(k-a-w)(w+1+a)}, \quad 1 \le v \le t-a-1, 0 \le w \le v \ . \tag{50}$$

Since, $l_{v,0} = \frac{c^v |\alpha|^v}{a!}$, therefore, $\frac{l_{v+1,0}}{\sum_{w=0}^{v} l_{vw}} \le \frac{l_{v+1,0}}{l_{v0}} \le c|\alpha|$. Therefore, for $1 \le v \le t-a-1$,

$$\frac{K_{v+1}}{2K_v} = \frac{\sum_{w=0}^{v+1} l_{v+1,w}}{2\sum_{w=0}^{v} l_{vw}} \le \frac{l_{v+1,0}}{l_{v_0}} + \max_{w=0}^{v} \left(\frac{l_{v+1,w+1}}{l_{vw}} \right) \le 2c|\alpha|, \text{ by Eqn. (50).}$$

or,

$$\frac{K_{v+1}}{K_v} \le \frac{\sum_{w=0}^{v+1} l_{v+1,w}}{\sum_{w=0}^{v} l_{vw}} \le 4c|\alpha| \le \frac{4(t-a)}{(k-a)(1-\frac{1}{25p})25p} \le \frac{1}{25p-1} = \frac{1}{49} \ .$$

$$\begin{aligned}
L &\le (1-|\alpha|)^{2p-2a} \beta^a t^{\underline{a}} \left(\frac{p^{\underline{a}}}{k^{\underline{a}}} \right)^2 \left(\frac{1}{a!} + \sum_{v=1}^{t-a} K_v \right) \\
&\le (1-|\alpha|)^{2p-2a} \beta^a t^{\underline{a}} \left(\frac{p^{\underline{a}}}{k^{\underline{a}}} \right)^2 \left(\frac{1}{a!} + \frac{49 K_1}{48} \right) \\
&= (1-|\alpha|)^{2p-2a} \beta^a t^{\underline{a}} \left(\frac{p^{\underline{a}}}{k^{\underline{a}}} \right)^2 \left(\frac{1}{a!} + \frac{49}{48} \left(\frac{c|\alpha|}{a!} + \frac{c\beta'|\alpha|}{(k-a)(a+1)!} \right) \right) \\
&\le (1-|\alpha|)^{2p-2a} \beta^a t^{\underline{a}} \left(\frac{p^{\underline{a}}}{k^{\underline{a}}} \right)^2 \left(\frac{1}{a!} \right) (1.006) \\
&\le (1.006)(1-|\alpha|)^{-2} \left(\frac{t^{\underline{a}}}{k^{\underline{a}}} \right) \left(\frac{\beta^a p^{\underline{a}}}{k^{\underline{a}}} \right) \left(\frac{p^{\underline{a}}}{a!} \right) \\
&\le (0.2625) \left(\frac{p^2 \beta}{k^a} \right) \left(\frac{1}{(2)(25)^2 p} \right)^{(a-1)}
\end{aligned}$$

since, (i) $c = \frac{(t-a)}{(k-a)(1-|\alpha|)} \le \frac{t}{k(1-\frac{1}{50})} \le 0.256$, (ii) $\left(\frac{t^{\underline{a}}}{k^{\underline{a}}}\right) \le \left(\frac{t}{k}\right)^a \le \left(\frac{1}{4}\right)^a$, (iii) $a = \lfloor p \rfloor + 1$ and therefore, $p^{\underline{a}} \le a!$, (iv) $\beta p \le \frac{(2)p}{(25p)^2} \le \frac{2}{(25)^2 p}$ and so, $\frac{\beta^a p^{\underline{a}}}{k^{\underline{a}}} \le \left(\frac{\beta p}{k}\right)^a = (p^2\beta)\frac{(p\beta)^{a-1}}{pk^a} \le (p^2\beta)\left(\frac{2}{(25)^2 p}\right)^{(a-1)} \frac{1}{k^a}$.

Substituting in Eqn. (51), we have,

$$P_1 \le (0.3) \left(\frac{p^2 \beta}{k^a} \right) \left(\frac{1}{(2)(25)^2 p} \right)^{(a-1)}$$

□

B.5 Completing Variance calculation for Averaged Taylor Polynomial Estimator

Lemma 30. *Assume the premises of Lemma 22 and let $\mu = \lambda$. Let $y, y' \in Y$ be distinct. Then,*

$$\mathsf{Cov}\left(\vartheta_y, \vartheta_{y'}\right) \leq \left(\frac{(0.261)p^2}{k}\right)\mu^{2p-2}\eta^2 \ .$$

Proof. By Lemma 22,

$$\begin{aligned}\mathsf{Cov}\left(\vartheta_y, \vartheta_{y'}\right) &= \sum_{v=1}^{k} \gamma_v^2(\lambda)\eta^{2v}\mathsf{Pr}_{\pi_y,\pi_{y'}}\left[q^{vv}_{yy'\pi_y\pi_{y'}} = v\right] \\ &= \sum_{v=1}^{k}\binom{p}{v}^2\lambda^{2(p-v)}\eta^{2v}\left(\frac{\binom{t}{v}}{\binom{k}{v}^2}\right)\end{aligned}$$

Taking the ratio of the $v+1$st term and the vth term of the summation above, we obtain,

$$\left(\frac{(p-v)^2}{(v+1)^2}\right)\left(\frac{\eta^2}{\lambda^2}\right)\left(\frac{(t-v)}{v+1}\right)\left(\frac{(v+1)}{k-v}\right)^2 \leq \left(\frac{(p-1)}{2}\right)\left(\frac{2}{(25p)^2}\right) \leq \frac{1}{2500}$$

Therefore,

$$\mathsf{Cov}\left(\vartheta_y, \vartheta_{y'}\right) \leq p^2\lambda^{2p-2}\eta^2\left(\frac{t}{k^2}\right)\left(1+\frac{1}{2499}\right) \leq \left(\frac{(0.251)p^2}{k}\right)\lambda^{2p-2}\eta^2 \ .$$

since, $\frac{t}{k} \leq \frac{1}{4}$. □

Lemma 31. *Assume the premises of Lemma 22. Let $y, y' \in Y$ be distinct. Then,*

$$\mathsf{Cov}\left(\vartheta_y, \vartheta_{y'}\right) \leq \frac{0.276p^2\lambda^{2p}\beta}{k} \ .$$

Proof. Case 1: $\mu = \lambda$. By Lemma 30,

$$\mathsf{Cov}\left(\vartheta_y, \vartheta_{y'}\right) \leq \left(\frac{(0.251)p^2}{k}\right)\lambda^{2p-2}\eta^2 \ .$$

Case 2: $\mu \neq \lambda$. Adding the expressions for P_3, P_2 and P_1 respectively from Lemmas 27 to 29, we obtain,

$$\begin{aligned}P &\leq \frac{p^2\lambda^{2p}\beta}{k}\left((0.275) + \left(\frac{1}{1200} + \left(\frac{0.3}{k^{a-1}}\right)\left(\frac{1}{(2)(25)^2p}\right)^{(a-1)}\right)\mathbf{1}_{p \text{ non-integral}}\right) \\ &\leq \frac{0.276p^2\lambda^{2p}\beta}{k} \qquad (51)\end{aligned}$$

Therefore,

$$\begin{aligned}\mathsf{Cov}\left(\vartheta_y, \vartheta_{y'}\right) &\leq P_{yy'} + Q_{yy'}, \qquad \text{by Corollary 23} \\ &\leq \frac{0.276p^2\lambda^{2p}\beta}{k}, \qquad \text{by Eqn. (51) and Lemma 25}\end{aligned}$$

Thus, in all cases,

$$\mathsf{Cov}\left(\vartheta_y, \vartheta_{y'}\right) \leq \frac{0.276 p^2 \lambda^{2p} \beta}{k} \ .$$

□

Lemma 32. *Assume the premises of Lemma 22 and let $k \geq 1000$ and $n \geq 2$. Then,*

$$\mathsf{Var}\left[\bar{\vartheta}\right] \leq \left(\frac{(0.288)p^2}{k}\right) \mu^{2p-2} \eta^2 \ .$$

Proof.

$$\begin{aligned}
\mathsf{Var}\left[\bar{\vartheta}\right] &= \frac{1}{|Y|^2} \sum_{y \in Y} \mathsf{Var}\left[\vartheta_y\right] + \sum_{\substack{y \neq y' \\ y, y' \in Y}} \mathsf{Cov}\left(\vartheta_y, \vartheta_{y'}\right) \\
&\leq \left(\frac{1}{|Y|^2}\right) |Y| (1.08) p^2 \mu^{2p-2} \eta^2 + \left(\frac{|Y|(|Y|-1)}{|Y|^2}\right) \left(\frac{(0.276) p^2 \lambda^{2p-2} \eta^2}{k}\right) \\
&= \left(\frac{1}{2^{0.08k}}\right) (1.08) p^2 \mu^{2p-2} \eta^2 + (0.276)(e^{1/25}) \left(\frac{p^2 \mu^{2p-2} \eta^2}{k}\right) \\
&\leq \left(\frac{(0.288)p^2}{k}\right) \mu^{2p-2} \eta^2 \qquad \text{for } k \geq 1000. \qquad (52)
\end{aligned}$$

The second step uses Corollary 3 and 31. □

C Proof that $\mathcal{G}$ holds with very high probability

C.1 Preliminaries and Auxiliary Events

The event GOODF$_2$. Using standard algorithms for estimating F_2 such as [1, 30], one can obtain an estimate $\tilde{F}_2$ satisfying $|\tilde{F}_2 - F_2| \leq \frac{0.001}{8p} F_2$, with probability $1 - n^{-25}$ using space $O(\log^2 n)$ bits. Then, $\hat{F}_2 = \left(1 - \frac{0.001}{8p}\right)^{-1} \tilde{F}_2$ satisfies $F_2 \leq \hat{F}_2 \leq \left(1 + \frac{0.001}{2p}\right) F_2$, which is the event GOODF$_2$.

The event GOODEST essentially states that the CountSketch guarantees for accuracy of estimation holds for all items and at all levels.

Lemma 33. GOODEST *holds with probability* $1 - n^{-23}$.

Proof. By guarantees of CountSketch structure [12] using tables with $16C_l$ buckets and $s = 8k = (8)(1000)(\log n)$ tables with independent hash functions, we have, $|\hat{f}_{il} - f_i| \leq \left(F_2^{\text{res}}\left(C_l, l\right)/C_l\right)^{1/2}$ with probability $1 - n^{-25}$. Using union bound to add the error probability over the levels $L = O(\log n)$ and $i \in [n]$, we obtain that GOODEST holds except with probability $n^{-25}(L)(n) \leq n^{-23}$. □

The above events comprising $\mathcal{G}$ will be shown to hold with probability $1 - n^{-\Omega(1)}$. In order to do so, we define a few auxiliary events.

Auxiliary Events

For $l \in \{0\} \cup [L]$ and $q \geq 1$, define the random variable

$$H_{lq} = \sum_{1 \leq \text{rank}(i) \leq 2^l q} y_{il} \text{ and } U_{lq} = \sum_{\text{rank}(i) > 2^l q} f_i^2 \cdot y_{il}$$

where, for $i \in [n]$, y_{il} is an indicator variable that is 1 if $i \in \mathcal{S}_l$ and is 0 otherwise. For $l \in \{0\} \cup [L]$, define two auxiliary events parameterized by a parameter q, as follows.

$$\begin{aligned} \textsc{small-h}(l,q) &\equiv H_{lq} \leq 2q, \quad \text{and} \\ \textsc{small-u}(l,q) &\equiv U_{l,q} \leq \frac{1.5 F_2^{\text{res}}\left(2^{l-1}q\right)}{2^{l-1}} \ . \end{aligned}$$

C.2 Proof that space parameter C_l is polynomial sized

We will now show that $C_l = n^{\Omega(1)}$ for each $l \in \{0\} \cup [L]$. This would also imply that $B_l = C_l(27p)^{-2} = n^{\Omega(1)}$ for each$l \in \{0\} \cup [L]$.

Lemma 34. *Assume the parameter values given in Figure 2. Then for $p > 2$, $C_L \geq n^{\Omega(1)}$.*

Proof. Since $L = \lceil \log_{2\alpha}(n/C) \rceil$,

$$\begin{aligned} C_L = 4\alpha^L C &\geq (4\alpha)\alpha^{\log_{2\alpha}(n/C)} C = \frac{(4\alpha)(2\alpha)^{\log_{2\alpha}(n/C)} C}{2^{\log_{2\alpha}(n/C)}} \\ &= \frac{4\alpha n}{(2^{\log_2(n/C)})^{1/(\log_2(2\alpha))}} = \frac{4\alpha n}{(n/C)^{1/\log_2(2\alpha)}} \ . \end{aligned} \tag{53}$$

Let $\alpha = 1 - \gamma$. Then,

$$\log_2(2\alpha) = 1 + \log_2(\alpha) = 1 + \frac{\ln(\alpha)}{\ln 2} \geq 1 - \frac{2\gamma}{\ln(2)}$$

since, $\gamma < 1/2$. Hence,

$$\frac{1}{\log_2(2\alpha)} = \frac{1}{(1 - 2\gamma/\ln(2))} \leq 1 + \frac{4\gamma}{\ln 2} \ .$$

Let $C = Kn^{1-2/p}$. Substituting in (53),

$$\begin{aligned} C_L &\geq \frac{4\alpha n}{(n/C)^{1/\log_2(2\alpha)}} \\ &\geq \frac{4\alpha n}{(n/C)^{1+4\gamma/\ln(2)}} \\ &= 4\alpha C (n/C)^{-4\gamma/\ln(2)} \\ &= 4\alpha K n^{1-2/p} \cdot (K^{-1} n^{2/p})^{-4\gamma/\ln(2)} \\ &= 4\alpha K \cdot K' \cdot n^{1-2/p-(2/p)(4\gamma/\ln(2))} \end{aligned} \tag{54}$$

where, $K' = K^{4\gamma/\ln(2)}$.

Since, $\alpha = 1-(1-2/p)\nu$, $\gamma = 1-\alpha = (1-2/p)\nu$. The exponent of n in (54) is

$$\begin{aligned} 1-2/p-(2/p)(4\gamma/\ln(2) &= 1-2/p-(2/p)\,(1-2/p)\left(\frac{4\nu}{\ln(2)}\right) \\ &= (1-2/p)\left(1-(2/p)\left(\frac{4\nu}{\ln 2}\right)\right) \end{aligned}$$

which is a positive constant for all $p > 2$ and $\nu < (\ln 2)/4$. Thus, $C_L = n^{\Omega(1)}$. □

Remark. This is the only place where the fact $p > 2$ is explicitly used. If $p = 2$, then, C_L would be $\Theta(\epsilon^{-2})$, and L would be $\log_2(n\epsilon^2) + O(1)$. The analysis would work, although the space bound would increase by a factor of $O(\log(n\epsilon^2))$.

C.3 Application of Chernoff-Hoeffding bounds for Limited Independence

We will use the following version of Chernoff-Hoeffding bounds for limited independence, specifically, Theorem 2.5 (II a) from [28].

Theorem 35 ([28])**.** *Let $X_1, X_2, \ldots, X_n$ be d-wise independent random variables with support in $[0,1]$. Let $X = X_1 + \ldots + X_n$, with $\mathbb{E}[X] = \mu$. Then, for $\delta \geq 1$ and $d \leq \lceil \delta\mu e^{-1/3} \rceil$, $\mathsf{Pr}[|X-\mu| \geq \delta\mu] \leq e^{-\lfloor d/2 \rfloor}$.*

The following lemma is shown whose proof is given later in this section.

Lemma 36. *Suppose $d \leq \lfloor qe^{-1/3} \rfloor$. Then, for $l \in \{0\} \cup [L]$ the following hold,*

1) $\mathsf{Pr}[\textsc{small-h}(l,q)] \geq 1 - e^{-\lfloor d/2 \rfloor}$, *and,*

2) *either $U_{lq} = 0$ or $\mathsf{Pr}[\textsc{small-u}(l,q)] \geq 1 - e^{-\lfloor d/2 \rfloor}$.*

Lemma 34 shows that $C_L = n^{\Omega(1)}$. This implies that $B_L = \bar{\epsilon}^2 n^{\Omega(1)} = n^{\Omega(1)}$ since $\bar{\epsilon} = 1/(27p)$. Therefore, $C_l > B_l \geq B_L = n^{\Omega(1)}$ for all $l \in \{0\} \cup [L]$. Hence we can use Lemma 36 and the union bound over $l \in \{0\} \cup [L]$ to show that the following events hold with probability $1 - Le^{-\lfloor d/2 \rfloor} = 1 - Le^{-\Omega(\log n)} \geq 1 - n^{-24}$, for suitable choice of the constant.

$$\begin{aligned} &(a)\ \wedge_{l \in \{0\}\cup[L]}\ \textsc{small-h}(l, C_l), \quad (b) \quad \wedge_{l \in \{0\}\cup[L]}\ \textsc{small-h}(H, C_l/2), \quad \text{and} \\ &(c)\ \wedge_{l \in \{0\}\cup[L]}\ \textsc{small-h}(l, \lceil \alpha^l B_l/(1-2\bar{\epsilon})^2 \rceil \end{aligned}$$

We now prove Lemma 36.

Proof of Lemma 36. For any fixed l, y_{il} is an indicator variable that is 1 iff $g_1(i) = g_2(i) = \ldots = g_l(i) = 1$. Since the g_l's are drawn independently from d-wise independent hash family, the y_{il}'s are d-wise independent.

By definition, $H_{lq} = \sum_{1 \leq rank(i) \leq 2^l q} y_{il}$ is the number of items with rank $2^l q$ or less that have hashed to level l. Since, $\mathsf{Pr}[y_{il}] = 1/2^l$, we have, $\mathbb{E}[H_{lq}] = 2^l q \cdot \frac{1}{2^l} = q$. Therefore,

$$\mathsf{Pr}[H_{lq} > 2q] \leq \mathsf{Pr}[|H_{lq} - q| > q] \leq e^{-\lfloor d/2 \rfloor}$$

by using Theorem 35 and assuming $d \leq qe^{-1/3}$.

We now prove the bound on U_{lq}. By definition, $U_{lq} = \sum_{\text{rank}(i)>2^l q} f_i^2 y_{il}$. Taking expectation, $\mathbb{E}[U_{lq}] = \sum_{\text{rank}(i)>2^l q} f_i^2/2^l = F_2^{\text{res}}\left(2^l q\right)/2^l$.

Since, $|f_{\text{rank}(2^l q)}| \leq |f_{\text{rank}(j)}|$ for each $j \in \{2^{l-1}q+1, \ldots, 2^l q\}$, it follows that $f^2_{\text{rank}(2^l q)} \leq F_2^{\text{res}}\left(2^{l-1}q\right)/(2^{l-1}q)$.

Case 1: Suppose $F_2^{res}\left(2^{l-1}q\right) > 0$. Define a scaled down variable U'_{lq} as follows.

$$U'_{lq} = \sum_{\text{rank}(i)>2^l q} \frac{f_i^2}{F_2^{\text{res}}\left(2^{l-1}q\right)/(2^{l-1}q)} \cdot y_{il} = \frac{(2^{l-1}q)U_{lq}}{F_2^{\text{res}}\left(2^{l-1}q\right)} \ .$$

By the above argument, the multiplier $f_i^2/(F_2^{\text{res}}\left(2^{l-1}q\right)/(2^{l-1}q)) \leq 1$. Since y_{il} are indicator variables, U'_{lq} is the sum of d-wise independent variables with support in the interval $[0,1]$.

Taking expectation,

$$\mathbb{E}\big[U'_{lq}\big] = \frac{(2^{l-1}q)\mathbb{E}[U_{lq}]}{F_2^{\text{res}}\left(2^{l-1}q\right)} = \frac{(2^{l-1}q)}{F_2^{\text{res}}\left(2^{l-1}q\right)} \cdot \frac{F_2^{\text{res}}\left(2^l q\right)}{2^l} \leq \frac{q}{2} \ .$$

By Theorem 35, we obtain,

$$\mathsf{Pr}\big[U'_{lq} > \mathbb{E}\big[U'_{lq}\big] + q\big] \leq \mathsf{Pr}\big[\big|U'_{lq} - \mathbb{E}\big[U'_{lq}\big]\big| > q\big] \leq e^{-\lfloor d/2 \rfloor} \ .$$

provided, $d \leq \lceil qe^{-1/3} \rceil$, which is assumed.

The event $U'_{lq} > \mathbb{E}\big[U'_{lq}\big] + q$ may be equivalently written (by rescaling) as $U_{lq} > \mathbb{E}\big[U_{lq}\big] + \frac{qF_2^{\text{res}}\left(2^{l-1}q\right)}{2^{l-1}q}$, which is the same as $U_{lq} > \frac{F_2^{\text{res}}\left(2^l q\right)}{2^l} + \frac{F_2^{\text{res}}\left(2^{l-1}q\right)}{2^{l-1}}$. This in turn is implied by the event $U_{lq} > \frac{1.5F_2^{\text{res}}\left(2^{l-1}q\right)}{2^{l-1}}$.

Therefore,

$$\mathsf{Pr}\Big[U_{lq} > \frac{1.5F_2^{\text{res}}\left(2^{l-1}q\right)}{2^{l-1}}\Big] \leq \mathsf{Pr}\left[U'_{lq} > \mathbb{E}\big[U'_{lq}\big] + q\right] \leq e^{-\lfloor d/2 \rfloor}$$

Case 2: $F_2^{res}\left(2^{l-1}q\right) = 0$. Then, $U_{lq} = 0$. □

Lemma 37. *$\forall l \in \{0\} \cup [L]$, SMALL-H$(l, C_l)$, SMALL-H$(l, \lceil B_l/(1-2\bar{\epsilon})^2 \rceil)$ and SMALL-U(l, C_l) hold simultaneously with probability $1 - O(n^{-25})$.*

Proof. From Lemma 36, SMALL-H(l, C_l) and SMALL-U(l, C_l) each holds with probability $e^{-\min(\lfloor d/2 \rfloor, C_l e^{-1/3}/2)}$. Similarly, SMALL-H$(l, \lceil B_l/(1-2\bar{\epsilon})^2 \rceil)$ holds with probability $e^{-\min\left(\lfloor d/2 \rfloor, \left(B_l/(1-2\bar{\epsilon})^2\right)e^{-1/3}/2\right)}$.

From Lemma 34, we have, $C_L \geq n^{\Omega(1)}$, and hence, $C_l \geq C_L \geq n^{\Omega(1)}$ for each $l \in \{0\} \cup [L]$. Hence, $d = O(\log n) = o(C_L) = O(C_l)$ for each l. The failure probability is therefore $e^{-d/2}$, since $\bar{\epsilon} = 1/(27p)$, $B_l = \bar{\epsilon}^2 C_l = n^{\Omega(1)}$ and therefore, $d = o(B_l)$, for each l.

Taking union bounds over the $O(\log n)$ values of l, the three events hold simultaneously except with probability $(L+1)(3)e^{-d/2} \leq (L+1)(3)e^{-50(\log n)/2} = o(n^{-24})$.

□

C.4 Proof that SMALLRES, ACCUEST, GOODL, SMALLHH hold with very high probability

Lemma 38. *Let $L = \lceil \log_{2\alpha}(n/C) \rceil$ and the hash functions $g_1, g_2, \ldots, g_L$ are drawn from d-wise independent family with $d = O(\log n)$ and even. Suppose* SMALL-H(l, C_l) *and* SMALL-U(l, C_l) *holds for each $l \in \{0\} \cup [L]$. Then,* SMALLRES *holds.*

Proof. We first show that SMALLRES$_l \equiv F_2^{\text{res}}(2C_l, l) \leq 1.5F_2^{\text{res}}\left((2\alpha)^l C\right)/2^{l-1}$ is implied by SMALL-H (l, C_l) and SMALL-U(l, C_l).

If SMALL-H(l, C_l) holds, then, $H_{l,C_l} \leq 2C_l$, that is, $\sum_{1 \leq \text{rank}(i) \leq 2^l C_l} y_{il} \leq 2C_l$. Hence,

$$F_2^{\text{res}}(2C_l, l) \leq \sum_{\text{rank}(i) > 2^l C_l} f_i^2 y_{il} = U_{l,C_l} \leq \frac{1.5F_2^{\text{res}}\left(2^{l-1}C_l\right)}{2^{l-1}}$$

where the last inequality follows since SMALL-U(l, C_l) holds.
Further, $2^{l-1}C_l = 2^{l-1}(4\alpha^l C) \geq 2(2\alpha)^l C$, since, $0 < \alpha < 1$. Thus,

$$F_2^{\text{res}}(2C_l, l) \leq \frac{(1.5)F_2^{\text{res}}\left(2(2\alpha)^l C_l\right)}{2^{l-1}} \ .$$

Hence SMALLRES$_l$ holds, for each $l \in \{0\} \cup [L]$, or equivalently, SMALLRES holds. □

Lemma 39. GOODEST $\wedge$ SMALLRES *imply* ACCUEST.

Proof. Fix $i \in [n]$ and $l \in \{0\} \cup [L]$. By construction, $C_l = 4\alpha^l C$. Thus,

$$|\hat{f}_{il} - f_i|^2 \leq \frac{F_2^{\text{res}}(C_l, l)}{C_l} \leq \frac{1.5F_2^{\text{res}}\left(2(2\alpha)^l C\right)}{2^{l-1}(4\alpha^l C)} \leq \frac{F_2^{\text{res}}\left((2\alpha)^l C\right)}{2(2\alpha)^l C}$$

where the first step follows from GOODEST and the second step follows from SMALLRES. □

We now show that the HH_L structure discovers all items and their exact frequencies that map to level L (with high probability).

Lemma 40. *For $L = \lceil \log_{2\alpha} \frac{n}{C} \rceil$ and assuming* SMALL-H(L, C_L) *and* GOODEST$_L$ *holds, the frequencies of all the items in $\mathcal{S}_L$ are discovered without error using HH_L. That is,* SMALL-H$(L, C_L) \wedge$ GOODEST$_L$ *implies* GOODFINALLEVEL.

Proof. Let $L = \lceil \log_{2\alpha}(n/C) \rceil$. Then,

$$2^L(C_L/2) = 2^L(4\alpha^L C/2) = 2(2\alpha)^L C \geq 2(n/C)C = 2n \ .$$

By definition, $H_{L,C_L/2} = \sum_{1 \leq \text{rank}(i) \leq 2^L(C_L/2)} y_{il}$ counts the number of items that map to level L with ranks in $1, 2, \ldots, 2^L(C_L/2)$. But $2^L(C_L/2) > n$. Hence, $H_{L,C_L/2}$ is the number of items that map to level l. Since, SMALL-H$(L, C_L/2)$ holds, $H_{L,C_L/2} \leq C_L$. Hence, $F_2^{\text{res}}(C_L, L) = 0$. By GOODEST$_L$, $|\hat{f}_{iL} - f_i| \leq \left(F_2^{\text{res}}(C_L, L)/C_L\right)^{1/2} = 0$. Thus if $i \in \mathcal{S}_L$ then $\hat{f}_{iL} = f_i$. □

Remark 1. Lemma 40 can be proved as an implication of the event SMALL-H(l, C_L) by using an ℓ_2/ℓ_1-compressed sensing recovery procedure as in [9, 14].

Remark 2. In the turnstile streaming model assumed, we say that i appears in the stream iff $|f_i| \geq 1$. By Lemma 40, the frequencies of all items are discovered exactly. Hence items with non-zero frequencies, that is, those with $|f_i| \geq 1$ would satisfy $|\hat{f}_{iL}| = |f_i| > 1/2 = Q_L$ and thus would qualify the criterion of being discovered at level L. All other items would satisfy $|\hat{f}_{iL}| = 0$ and will not be discovered at level L.

At each level l, the algorithm finds the top-C_l items by absolute values of estimated frequencies. A heavy-hitter at a level l is however defined as an item whose estimated frequency crosses the threshold Q_l. The event SMALLHH$_l$ states that the heavy-hitters at a level l are always among the top-C_l items by absolute estimated frequencies.

Lemma 41. *Suppose* SMALL-H$(l, \lceil B_l/(1-2\bar{\epsilon})^2 \rceil)$ *holds for each* $l \in \{0\} \cup [L-1]$ *and suppose* ACCUEST *holds. Then,* SMALLHH *holds.*

Proof. Let H'_l denote the set of items that are discovered as heavy-hitters at level l, that is, $H'_l = \{i \in \mathcal{S}_l \mid |\hat{f}_i| \geq Q_l\}$, where, $Q_l = T_l(1-\bar{\epsilon})\}$. By ACCUEST and since $\bar{\epsilon} = (B/C)^{1/2}$, we obtain

$$|\hat{f}_{il} - f_i| \leq \left(\frac{F_2^{\text{res}}\left((2\alpha)^l C\right)}{2(2\alpha)^l C} \right)^{1/2} \leq \frac{\bar{\epsilon}}{\sqrt{2}} \left(\frac{F_2}{(2\alpha)^l B} \right)^{1/2} \quad .$$

Suppose $i \in H'_l$. Then,

$$|f_i| \geq Q_l - \frac{\bar{\epsilon}}{\sqrt{2}} \left(\frac{F_2}{(2\alpha)^l B} \right)^{1/2} \geq T_l(1-\bar{\epsilon}) - T_l(\bar{\epsilon}/\sqrt{2}) \geq T_l\left(1 - 2\bar{\epsilon}\right) \quad .$$

since, $T_l = (\hat{F}_2/((2\alpha)^l B))^{1/2} \geq (F_2/((2\alpha)^l B))^{1/2}$.
Therefore,

$$\text{rank}(i) \leq \frac{F_2}{|f_i|^2} \leq \frac{F_2}{(T_l(1-2\bar{\epsilon}))^2} = \frac{F_2(2\alpha)^l B}{\hat{F}_2(1-2\bar{\epsilon})^2} \leq \frac{2^l B_l}{(1-2\bar{\epsilon})^2}$$

Hence $H'_l \subset H_{lq}$, where we let $q = B_l/(1-2\bar{\epsilon})^2$.
Since SMALL-H(l, q) holds, $H_{lq} \leq 2q$. Further, since, $H'_l \subset H_{lq}$, therefore, $|H'_l| \leq 2q = 2B_l/(1-2\bar{\epsilon})^2 \leq C_l$, since, by choice of parameters, $\bar{\epsilon} = (B_l/C_l)^{1/2} = 1/(27p)$ and $p \geq 1$.
By construction, H'_l is the set of items whose estimated frequencies are at least Q_l. Hence,

$$H'_l = \widehat{\text{TOPK}}(|H'_l|) \subset \widehat{\text{TOPK}}(C_l) \quad .$$

□

C.5 Proof that NOCOLLISION holds with very high probability

Lemma 42. *If* $t \geq 6$ *and* $s = \Theta(\log n)$*, then,* NOCOLL *holds with probability at least* $1 - n^{-150}$.

Proof. Assume full independence of hash functions. For $i \in \widehat{\textsc{Topk}}_l(C_l)$ and $l \in [2s]$, let $w_{ijl} = 1$ if i collides with some other item in $\widehat{\textsc{Topk}}_l(C_l)$ in the jth table of the TPEST structure at level l. Since, each table at level $l \in \{0\} \cup [L-1]$ has $16C_l$ buckets, therefore,

$$q = \mathsf{Pr}\left[w_{ijl} = 1\right] = 1 - \left(1 - \frac{1}{16C_l}\right)^{C_l - 1} \leq 1/16 \ .$$

Let $W_{il} = \sum_{j=1}^{2s}(1 - w_{ijl})$ be the number of tables where i does not collide with any other item of $\widehat{\textsc{Topk}}_l(C_l)$. Then, $\mathbb{E}\big[W_{il}\big] \geq (1-q)(2s) \geq (15/8)s$. By Chernoff's bounds,

$$\begin{aligned}\mathsf{Pr}\left[W_{il} \geq s\right] &\geq 1 - \ \exp\left\{-(15/8)s(7/15)^2/2\right\} \geq 1 - e^{-0.2s} \\ &= 1 - e^{-(0.2)(8)(100)\log(n)} = 1 - n^{-160}\end{aligned}$$

since, $s = 8k = 8(100\log(n))$.
By union bound,

$$\mathsf{Pr}\left[\forall i \in \widehat{\textsc{Topk}}_l(C_l)\,(W_{il} \geq s)\right] \geq 1 - C_l e^{-0.2s} \geq 1 - n^{-150} \ .$$

Assuming t-wise independence of the hash family from which the h_{lj}'s are drawn, denote $q'_t = \mathsf{Pr}_t\big[w_{ijl} = 1\big]$, where the subscript t denotes t-wise independence. Let $u_{ikjl} = 1$ if i and k collide under hash function h_{lj} for the jth hash table in the structure $\textsc{tpest}_l$. Let $S_{li} = \overline{\textsc{Topk}}(C_l) \setminus \{i\}$. Then, by inclusion-exclusion,

$$\begin{aligned}1 - q = \mathsf{Pr}\big[w_{ijl} = 0\big] &= 1 - \mathsf{Pr}\big[w_{ijl} = 1\big] = 1 - \mathsf{Pr}\left[\bigvee_{k \in S_{li}} (u_{ikjl} = 1)\right] \\ &= 1 - \sum_{r=1}^{|S_{li}|}(-1)^{r-1} \sum_{\{k_1,k_2,\ldots,k_r\} \subset S_{li}} \mathsf{Pr}\big[u_{ik_1jl} = 1, u_{ik_2jl} = 1, \ldots, u_{ik_rjl} = 1\big] \qquad (55) \\ 1 - q'_t = \mathsf{Pr}_t\left[w_{ijl} = 0\right] &\\ = 1 - \sum_{r=1}^{|S_{li}|}(-1)^{r-1} &\sum_{\{k_1,k_2,\ldots,k_r\} \subset S_{li}} \mathsf{Pr}_t\big[u_{ik_1jl} = 1, u_{ik_2jl} = 1, \ldots, u_{ik_rjl} = 1\big] \qquad (56)\end{aligned}$$

Further, the sum of the tail starting from position $t+1$ to $|S_{li}|$ is, in absolute value, dominated by the tth term. Therefore, from (55), we have,

$$\begin{aligned}\Big|q - \sum_{r=1}^{t-1}(-1)^{r-1} &\sum_{\{k_1,k_2,\ldots,k_r\} \subset S_{li}} \mathsf{Pr}\big[u_{ik_1jl} = 1, u_{ik_2jl} = 1, \ldots, u_{ik_rjl} = 1\big]\Big| \\ &\leq \sum_{\{k_1,k_2,\ldots,k_t\} \subset S_{li}} \mathsf{Pr}\big[u_{ik_1jl} = 1, u_{ik_2jl} = 1, \ldots, u_{ik_rjl} = 1\big] \qquad (57)\end{aligned}$$

Similarly from (56), we have,

$$\begin{aligned}\Big|q' - \sum_{r=1}^{t-1}(-1)^{r-1} &\sum_{\{k_1,k_2,\ldots,k_r\} \subset S_{li}} \mathsf{Pr}_t\big[u_{ik_1jl} = 1, u_{ik_2jl} = 1, \ldots, u_{ik_rjl} = 1\big]\Big| \\ &\leq \sum_{\{k_1,k_2,\ldots,k_t\} \subset S_{li}} \mathsf{Pr}_t\big[u_{ik_1jl} = 1, u_{ik_2jl} = 1, \ldots, u_{ik_rjl} = 1\big] \qquad (58)\end{aligned}$$

By t-wise independence, the probability terms in the above expression are identical for $r = 1, \ldots, t$, that is, for any $1 \le k_1 < k_2 < \ldots < k_r \le n$ and $2 \le r \le t$.

$$\begin{aligned}&\mathsf{Pr}_t\big[u_{ik_1jl} = 1, u_{ik_2jl} = 1, \ldots, u_{ik_rjl} = 1\big]\\ &= \mathsf{Pr}\big[u_{ik_1jl} = 1, u_{ik_2jl} = 1, \ldots, u_{ik_rjl} = 1\big]\end{aligned}$$

Therefore, by triangle inequality,

$$|q - q'| \le 2 \sum_{j_1 < j_2 < \ldots < j_t} \mathsf{Pr}_t\left[u_{ik_1jl} = 1, u_{ik_2jl} = 1, u_{ik_tjl} = 1\right] \tag{59}$$

Since there are $16C_l$ buckets in the TPEST structure at level l, we have, $\mathsf{Pr}\big[u_{ik_rjl} = 1\big] = 1/(16C_l)$. Substituting in (59),

$$\begin{aligned}\big|q - q'_t\big| &\le 2 \sum_{j_1 < j_2 < \ldots < j_t} \mathsf{Pr}_t\big[u_{ik_1jl} = 1, u_{ik_2jl} = 1, u_{ik_tjl} = 1\big]\\ &= 2\binom{|S_{li}|}{t}\left(\frac{1}{16C_l}\right)^t \le 2\binom{(C_1 - 1)}{t}(16C_l)^{-t} \le 2\left(\frac{C_l e}{16C_l t}\right)^t \le 2\left(\frac{e}{16t}\right)^t\end{aligned}$$

since, $|S_{li}| = C_l - 1$. For $t \ge 6$, $|q - q'_t| \le 2(32)^{-6} \le 2^{-29}$.

The above Chernoff's bound argument may be repeated using probability of success $1 - q'_t \ge 1 - q - 2^{-29}$, instead of $1 - q$. Hence, NOCOLL(H) holds except with probability n^{-150} by calculations similar to the previous one. □

C.6 Proof that $\mathcal{G}$ holds with very high probability

Restated Lemma (Restatement of Lemma 7)**.** $\mathsf{Pr}[\mathcal{G}] \ge 1 - O(n^{-24})$.

Proof. By adding the failure probabilities of all the events comprising $\mathcal{G}$ using Lemmas 33 through 37, the statement of the lemma follows. □

C.7 Technical fact

The following fact gives a bound on the difference between the unconditional probability of an event E and its probability conditioned on an event F. It essentially shows that if $\mathsf{Pr}[E] = 1/n^{O(1)}$, its probability is not significantly altered if it is conditioned by a very high probability event F, that is, $\mathsf{Pr}[F] = 1 - n^{-\Omega(1)}$.

Fact 43. *Let E and F be a pair of events such that $\mathsf{Pr}[F] > 0$. Then, $\big|\mathsf{Pr}[E \mid F] - \mathsf{Pr}[E]\big| \le 1 - \mathsf{Pr}[F]$.*

Proof of Fact 43. If $\mathsf{Pr}[F] = 1$, then $\mathsf{Pr}[E, F] = \mathsf{Pr}[E \cup F] - \mathsf{Pr}[E] - \mathsf{Pr}[F] = 1 - \mathsf{Pr}[E] - 1 = \mathsf{Pr}[E]$, and hence the statement holds. Otherwise,

$$\mathsf{Pr}[E] = \mathsf{Pr}[E \mid F]\,\mathsf{Pr}[F] + \mathsf{Pr}[E \mid \neg F]\,\mathsf{Pr}[\neg F]$$

Subtracting $\mathsf{Pr}[E \mid F]$ from both sides yields,

$$\begin{aligned}\mathsf{Pr}[E] - \mathsf{Pr}[E \mid F] &= \mathsf{Pr}[E \mid F]\,(\mathsf{Pr}[F] - 1) + \mathsf{Pr}[E \mid \neg F]\,\mathsf{Pr}[\neg F]\\ &= (-\mathsf{Pr}[E \mid F] + \mathsf{Pr}[E \mid \neg F])\,\mathsf{Pr}[\neg F]\end{aligned}$$

Taking absolute values and noting that $|-\mathsf{Pr}[E \mid F]+\mathsf{Pr}[E \mid \neg F]| \leq 1$, we have, $\left|\mathsf{Pr}[E]-\mathsf{Pr}[E \mid F]\right| \leq \mathsf{Pr}[\neg F]$.

□

The fact is used by letting $F = \mathcal{G}$. Then, for any event E, $\left|\mathsf{Pr}[E \mid \mathcal{G}] - \mathsf{Pr}[E]\right| \leq \mathsf{Pr}[\neg G] = O(n^{-24})$, by Lemma 7.

D Basic Sampling Properties of Geometric-Hss Algorithm

Preliminaries. The following lemma argues that the frequency ranges defining lmargin, mid and rmargin are non-empty intervals.

Lemma 44. *For $p \geq 2$ and for each $l \in \{0\} \cup [L]$, the frequency ranges that define lmargin(G_l), mid (G_l) and rmargin(G_l) are non-empty intervals.*

Proof. The statement of the lemma is obviously true from the definitions for lmargin(G_l) and rmargin(G_l).

For mid(G_l), the interval range is $[T_l(1+\bar{\epsilon}), T_{l-1}(1-2\bar{\epsilon}))$. This range is non-empty iff $T_{l-1}(1-2\bar{\epsilon}) > T_l(1+\bar{\epsilon})$, or, $T_{l-1}/T_l > (1+\bar{\epsilon})/(1-2\bar{\epsilon})$, or, $(2\alpha)^{1/2} > (1+1/(27p))/(1-2/(27p))$, which is true for $\alpha = 1 - 2(0.01)/p \geq 0.99$. □

Our analysis is conditioned on $\mathcal{G}$. Assuming $\mathcal{G}$ holds, the event ACCUEST holds, and therefore, the frequency estimation error by the HH_l structure is bounded as follows.

$$|\hat{f}_{il} - f_i| \leq \left(\frac{F_2^{\text{res}}\left((2\alpha)^l C\right)}{(2\alpha)^l C}\right)^{1/2} \leq \left(\frac{\hat{F}_2}{(2\alpha)^l C}\right)^{1/2} = \bar{\epsilon} T_l \ . \tag{60}$$

We first prove a property about the relation between the level at which an item is discovered and the group G_l to which an item belongs. This property is then used to a relation between the probabilities with which an item may belong to different sampled groups.

D.1 Properties concerning levels at which an item is discovered

Lemma 45. *The following properties hold conditional on $\mathcal{G}$.*

1) *Suppose $i \in$ lmargin(G_l) for some $0 \leq l \leq L-1$. Then, (a) $\mathsf{Pr}[l_d(i) \leq l-1 \mid \mathcal{G}] = 0$, and (b) the event $\{l_d(i) = l, \mathcal{G}\} \equiv \{i \in \mathcal{S}_l, \mathcal{G}\}$.*

2) *Suppose $i \in$ mid(G_l) for some $0 \leq l \leq L$. Then, (a) $\mathsf{Pr}[l_d(i) \leq l-1 \mid \mathcal{G}] = 0$, (b) the event $\{l_d(i) = l, \mathcal{G}\} \equiv \{i \in \mathcal{S}_l, \mathcal{G}\}$, and, (c) $\mathsf{Pr}\left[\hat{f}_{il} \geq T_l \mid i \in \mathcal{S}_l, \mathcal{G}\right] = 1$.*

3) *Suppose $i \in$ rmargin(G_l) for some $2 \leq l \leq L$. Then, (a) $\mathsf{Pr}[l_d(i) \leq l-2 \mid \mathcal{G}] = 0$, (b) $\{i \in \mathcal{S}_l, \mathcal{G}\}$ implies $\{|\hat{f}_{il}| \geq T_l\}$, and (c) $\mathsf{Pr}[l_d(i) = l \mid l_d(i) \neq l-1, \mathcal{G}] = \mathsf{Pr}[i \in \mathcal{S}_l \mid l_d(i) \neq l-1, \mathcal{G}]$.*

Proof of Lemma 45. Since, ACCUEST holds as a sub-event of $\mathcal{G}$, we have, $|\hat{f}_{il} - f_i| \leq \bar{\epsilon} T_l$, by Eqn. (60). Also, $Q_l = T_l(1-\bar{\epsilon})$. All statements below are conditional on $\mathcal{G}$.

Case: $i \in \mathit{lmargin}(G_l) \cup \mathit{mid}(G_l)$, $l \geq 1$. Then, $T_l + \bar{\epsilon}T_l \leq |f_i| < T_{l-1} - 2\bar{\epsilon}T_{l-1}$. Therefore for $r \leq l-1$,

$$|\hat{f}_{ir}| \leq |f_i| + \bar{\epsilon}T_r < T_{l-1} - 2\bar{\epsilon}T_{l-1} + \bar{\epsilon}T_r \leq T_r - 2\bar{\epsilon}T_r + \bar{\epsilon}T_r = T_r - \bar{\epsilon}T_r = Q_r \ .$$

Hence, $\mathsf{Pr}\left[l_d(i) \leq l-1 \mid \mathcal{G}\right] = 0$.

Further, if $i \in \mathcal{S}_l$ and $i \in \mathrm{lmargin}(G_l) \cup \mathrm{mid}(G_l)$, then, $|\hat{f}_{il}| \geq |f_i| - \bar{\epsilon}T_l \geq T_l - \bar{\epsilon}T_l = Q_l$ and so i is discovered at level l, if i has not been discovered at an earlier level. However, part(a) states that i cannot be discovered at levels $l-1$ or less. Hence i is discovered at level l. Thus, conditional upon $\mathcal{G}$, if $i \in \mathcal{S}_l$, then, $l_d(i) = l$. Conversely, if $i \notin \mathcal{S}_l$, then $l_d(i) \neq l$. Hence, the events $\{i \in \mathcal{S}_l\}$ and $\{l_d(i) = l\}$ are equivalent, conditional on $\mathcal{G}$. This proves parts 1(b) and 2(b).

Case: $i \in \mathit{mid}(G_l)$. If $i \in \mathcal{S}_l$, then, $|\hat{f}_{il}| \geq |f_i| - \bar{\epsilon}T_l \geq T_l + \bar{\epsilon}T_l - \bar{\epsilon}T_l = T_l$. This proves part 2(c).

Case: $i \in \mathit{rmargin}(G_l)$. Then, $|f_i| < T_{l-1}$. Let $r \leq l-2$. Then,

$$|\hat{f}_{ir}| \leq |f_i| + \bar{\epsilon}T_r < T_{l-1} + \bar{\epsilon}T_r < T_r - \bar{\epsilon}T_r = Q_r$$

where, the last inequality $T_{l-1} + \bar{\epsilon}T_r < T_r - \bar{\epsilon}T_r$ follows since, it is equivalent to $\frac{T_{l-1}}{T_{l-2}} < (1-2\bar{\epsilon})$, which holds since, $\frac{T_{l-1}}{T_{l-2}} = \frac{1}{\sqrt{2\alpha}} \leq (0.72)$ and $(1-2\bar{\epsilon}) = 1 - \frac{2}{27p} \geq 0.96$. Hence $\mathsf{Pr}\left[l_d(i) \leq l-2 \mid G\right] = 0$.

We are given that $i \in \mathrm{rmargin}(G_l)$. Suppose that $i \in \mathcal{S}_l$. Then,

$$\begin{aligned} |\hat{f}_{il}| &\geq T_{l-1} - 2\bar{\epsilon}T_{l-1} - \bar{\epsilon}T_l = T_l(2\alpha)^{1/2} - 2(2\alpha)^{1/2}\bar{\epsilon}T_l - \bar{\epsilon}T_l \\ &\geq T_l\left(1.40(1-(2)(0.04)) - (0.04)\right) = 1.248T_l > T_l \end{aligned} \tag{61}$$

Hence, $\mathsf{Pr}\left[|\hat{f}_{il}| \geq T_l \mid i \in \mathcal{S}_l, G\right] = 1$, and therefore, by Eqn. (61) $\mathsf{Pr}\left[l_d(i) \in \{l-1, l\} \mid i \in \mathcal{S}_l, G\right] = 1$. This proves part 2(b).

Since, $l_d(i) > l$ implies $i \in \mathcal{S}_l$, we have,

$$\begin{aligned} \mathsf{Pr}\left[l_d(i) > l \mid \mathcal{G}\right] &= \mathsf{Pr}\left[l_d(i) > l, i \in \mathcal{S}_l \mid \mathcal{G}\right] \\ &= \mathsf{Pr}\left[l_d(i) > l \mid i \in \mathcal{S}_l, \mathcal{G}\right] \cdot \mathsf{Pr}\left[i \in \mathcal{S}_l \mid \mathcal{G}\right] \\ &\leq \left(1 - \mathsf{Pr}\left[l_d(i) \in \{l-1, l\} \mid i \in \mathcal{S}_l, \mathcal{G}\right]\right) \cdot \mathsf{Pr}\left[i \in \mathcal{S}_l \mid \mathcal{G}\right] \\ &= 0 \ . \end{aligned}$$

Hence,

$$\begin{aligned} \mathsf{Pr}\left[l_d(i) \neq l-1 \mid \mathcal{G}\right] &= \mathsf{Pr}\left[l_d(i) \leq l-2 \mid \mathcal{G}\right] + \mathsf{Pr}\left[l_d(i) = l \mid \mathcal{G}\right] + \mathsf{Pr}\left[l_d(i) > l \mid \mathcal{G}\right] \\ &= 0 + \mathsf{Pr}\left[l_d(i) = l \mid \mathcal{G}\right] + 0 \ . \end{aligned} \tag{62}$$

It follows that,

$$\mathsf{Pr}\left[l_d(i) = l \mid l_d(i) \neq l-1, \mathcal{G}\right] = \frac{\mathsf{Pr}\left[l_d(i) = l \mid \mathcal{G}\right]}{\mathsf{Pr}\left[l_d(i) \neq l-1 \mid \mathcal{G}\right]} = 1$$

by Eqn. (62). □

D.2 Probability of items belonging to sampled groups

Restated Lemma (Re-statement of Lemma 8.)**.** *Let $i \in G_l$.*

1) *Suppose $i \in mid(G_l)$. Then, (a) the event $\{i \in \bar{G}_l, \mathcal{G}\} \equiv \{i \in \mathcal{S}_l, \mathcal{G}\}$, (b) $2^l \mathsf{Pr}\left[i \in \bar{G}_l \mid \mathcal{G}\right] = 1 \pm 2^l n^{-c}$, and, (c) $\mathsf{Pr}\left[i \in \cup_{l' \neq l} \bar{G}_{l'} \mid \mathcal{G}\right] = 0$.*

2) *Suppose $i \in lmargin(G_l)$. Then, (a) $\mathsf{Pr}\left[i \in \cup_{l' \neq \{l, l+1\}} \bar{G}_{l'}\right] = 0$, (b) the event $\{i \in \bar{G}_l \cup \bar{G}_{l+1}, \mathcal{G}\} \equiv \{i \in \mathcal{S}_l, \mathcal{G}\}$, and (c) $2^{l+1} \mathsf{Pr}\left[i \in \bar{G}_{l+1} \mid \mathcal{G}\right] + 2^l \mathsf{Pr}\left[i \in \bar{G}_l \mid \mathcal{G}\right] = 1 \pm 2^l n^{-c}$.*

3) *Suppose $i \in rmargin(G_l)$. Then, (a) $\mathsf{Pr}\left[i \in \cup_{l' \neq \{l-1, l\}} \bar{G}_{l'}\right] = 0$, (b) the events $\{i \in \bar{G}_{l-1} \cup \bar{G}_l\} \subset \{i \in \mathcal{S}_l\}$, (c) $\{i \in \mathcal{S}_l, l_d(i) \neq l-1\} \subset \{i \in \bar{G}_l\}$, and, (d) $2^l \mathsf{Pr}\left[i \in \bar{G}_l \mid \mathcal{G}\right] + 2^{l-1} \mathsf{Pr}\left[i \in \bar{G}_{l-1} \mid \mathcal{G}\right] = 1 \pm O(2^l n^{-c})$.*

Proof of Lemma 8. Assume $\mathcal{G}$ holds for the arguments in this proof. Suppose $i \in \mathcal{S}_l$. Then $|\hat{f}_{il} - f_i| \leq T_l \bar{\epsilon}$.

Case: $i \in mid(G_l)$. Part 1 (b). Since $i \in \text{mid}(G_l)$, $|f_i| \geq T_l + \bar{\epsilon} T_l$. Conditional on $\mathcal{G}$, ACCUEST holds, and therefore,

$$|\hat{f}_{il}| \geq |f_i| - \bar{\epsilon} T_l \geq T_l + \bar{\epsilon} T_l - \bar{\epsilon} T_l = T_l \ .$$

Therefore,

$$\{i \in \mathcal{S}_l, \mathcal{G}\} \subset \{|\hat{f}_{il}| \geq T_l, \mathcal{G}\} \tag{63}$$

Then,

$$\begin{aligned}
\{i \in \bar{G}_l, \mathcal{G}\} &\equiv \{l_d(i) = l, |\hat{f}_{il}| \geq T_l, \mathcal{G}\} \\
&\equiv \{i \in \mathcal{S}_l, |\hat{f}_{il}| \geq T_l, \mathcal{G}\}, \quad \text{since, } \{l_d(i) = l, \mathcal{G}\} \equiv \{i \in \mathcal{S}_l, \mathcal{G}\}, \text{ Lemma 45, (2b)} \\
&\equiv \{i \in \mathcal{S}_l, \mathcal{G}\}, \qquad \text{by Eqn. (63).}
\end{aligned}$$

This proves part 1 (b).

Part 1 (a).

$$\mathsf{Pr}\left[i \in \bar{G}_l \mid \mathcal{G}\right] = \mathsf{Pr}\left[l_d(i) = l, |\hat{f}_{il}| \geq T_l \mid \mathcal{G}\right] + \mathsf{Pr}\left[l_d(i) = l-1, Q_l \leq |\hat{f}_{i,l-1}| < T_l, K_i = 1 \mid \mathcal{G}\right] \tag{64}$$

Denote by $\mathcal{E}_1$ the event $l_d(i) = l-1, Q_l \leq |\hat{f}_{i,l-1}| < T_l$ and by $\mathcal{E}_2$ the event $Q_l \leq |\hat{f}_{i,l-1}| < T_l$. Then,

$$\mathsf{Pr}\left[\mathbb{E}_1, K_i = 1 \mid \mathcal{G}\right] = \mathsf{Pr}\left[K_i = 1 \mid \mathbb{E}_1, \mathcal{G}\right] \cdot \mathsf{Pr}\left[\mathbb{E}_2 \mid l_d(i) = l-1, \mathcal{G}\right] \cdot \mathsf{Pr}\left[l_d(i) = l-1 \mid \mathcal{G}\right] = 0$$

since, $\mathsf{Pr}\left[l_d(i) = l-1 \mid \mathcal{G}\right] = 0$, by Lemma 45, part (2a). Substituting in Eqn. (64), we have,

$$\begin{aligned}
\mathsf{Pr}\left[i \in \bar{G}_l \mid \mathcal{G}\right] &= \mathsf{Pr}\left[l_d(i) = l, |\hat{f}_{il}| \geq T_l \mid \mathcal{G}\right] \\
&= \mathsf{Pr}\left[i \in \mathcal{S}_l, |\hat{f}_{il}| \geq T_l \mid \mathcal{G}\right], \text{ since, } \{l_d(i) = l, \mathcal{G}\} \equiv \{i \in \mathcal{S}_l, \mathcal{G}\}, \text{ Lemma 45, (2b)} \\
&= \mathsf{Pr}\left[i \in \mathcal{S}_l \mid \mathcal{G}\right], \qquad \text{by part 1 (a)} \\
&= 2^{-l} \pm n^{-c}, \qquad \text{by Fact 43.}
\end{aligned}$$

Multiplying by 2^l and transposing, we have $2^l \mathsf{Pr}\left[i \in \bar{G}_l \mid \mathcal{G}\right] \in 1 \pm n^{-c} \cdot 2^l$, as claimed in part 1(a).

Part 1(c). We have by ACCUEST that for any $0 \leq r \leq l-1$,

$$|\hat{f}_{i,r}| < T_{l-1} - 2\bar{\epsilon}T_{l-1} + \bar{\epsilon}T_r \leq T_{l-1}\,(1-\bar{\epsilon}) = Q_{l-1}$$

Hence, i cannot be in $\bar{G}_r$ for any $r \leq l-1$. We have by part (1a) that $\{i \in \bar{G}_l, \mathcal{G}\} \equiv \{i \in \mathcal{S}_l, \mathcal{G}\}$.

Let $i \in \bar{G}_r$ for some $r \geq l+1$. Since, for i to belong to $\bar{G}_r$, i must be in $\mathcal{S}_{r-1}$ and hence by the sub-sampling procedure, $i \in \mathcal{S}_l$. By part 1(a), $\{i \in \bar{G}_l, \mathcal{G}\} \equiv \{i \in \mathcal{S}_l, \mathcal{G}\}$, and therefore, $i \in \bar{G}_l$. Hence, $i \notin \bar{G}_r$, for any $r \geq l+1$. Thus,

$$\mathsf{Pr}\left[i \in \cup_{r \neq l} \bar{G}_r \mid \mathcal{G}\right] = 0 \ .$$

Case: $i \in$ *lmargin*(G_l). From Lemma 45, $l_d(i) \not< l$ and $l_d(i) = l$ iff $i \in \mathcal{S}_l$. Since $l_d(i) \not< l$, $i \notin \bar{G}_r$, for any $r < l$. Consider $r > l+1$. If $i \in \bar{G}_r$, then, $l_d(i) \geq r-1 \geq l+1$. Since, $i \in \mathcal{S}_{l_d(i)}$, and $l_d(i) \geq l+1$, it follows that $i \in \mathcal{S}_l$, by the sub-sampling procedure. However, by Lemma 45, part (1b), $\{l_d(i) = l, \mathcal{G}\} \equiv \{i \in \mathcal{S}_l, \mathcal{G}\}$. Hence, in this case, $l_d(i) = l$, contradicting the implication that $l_d(i) \geq l+1$. Thus,

$$\mathsf{Pr}\left[i \in \cup_{l' \notin \{l, l+1\}} \bar{G}_{l'}\right] = 0$$

proving part 2 (a).

Suppose $i \in \mathcal{S}_l$. Then, $l_d(i) = l$ and $\hat{f}_i = \hat{f}_{il}$. By construction,

$$\mathsf{Pr}\left[i \in \bar{G}_l \mid i \in \mathcal{S}_l, \mathcal{G}\right] = \mathsf{Pr}\left[|\hat{f}_{il}| \geq T_l \mid i \in \mathcal{S}_l, \mathcal{G}\right] = p_{il} \quad \text{(say.)} \tag{65}$$

Further,

$$\begin{aligned}\mathsf{Pr}\left[i \in \bar{G}_{l+1} \mid i \in \mathcal{S}_l, \mathcal{G}\right] &= \mathsf{Pr}\left[Q_l \leq |\hat{f}_{il}| < T_l, K_i = 1 \mid i \in \mathcal{S}_l, \mathcal{G}\right] \\ &\quad + \mathsf{Pr}\left[|\hat{f}_{il}| < Q_l, i \in \mathcal{S}_{l+1}, |\hat{f}_{i,l+1}| \geq T_{l+1} \mid i \in \mathcal{S}_l, \mathcal{G}\right]\end{aligned} \tag{66}$$

However, conditional on $\mathcal{G}$ and $i \in \mathcal{S}_l$, by Lemma 45, $|\hat{f}_{il}| \geq Q_l$. Hence, the second probability in the *RHS* of Eqn. (66) is 0. Therefore,

$$\begin{aligned}&\mathsf{Pr}\left[i \in \bar{G}_{l+1} \mid i \in \mathcal{S}_l, \mathcal{G}\right] \\ &= \mathsf{Pr}\left[Q_l \leq |\hat{f}_{il}| < T_l, K_i = 1 \mid i \in \mathcal{S}_l, \mathcal{G}\right] \\ &= \mathsf{Pr}\left[K_i = 1 \mid Q_l \leq |\hat{f}_{il}| < T_l, i \in \mathcal{S}_l, \mathcal{G}\right] \cdot \mathsf{Pr}\left[Q_l \leq |\hat{f}_{il}| < T_l \mid i \in \mathcal{S}_l, \mathcal{G}\right] \\ &= (1/2)\,(1-p_{il})\end{aligned} \tag{67}$$

since, (a) K_i is independent of all other random bits, and, (b) $\mathsf{Pr}\big[Q_l \leq |\hat{f}_{il}| < T_l \mid i \in \mathcal{S}_l, \mathcal{G}\big] + \mathsf{Pr}\left[|\hat{f}_{il}| \geq T_l \mid i \in \mathcal{S}_l, \mathcal{G}\right] = \mathsf{Pr}\left[|\hat{f}_{il}| \geq Q_l \mid i \in \mathcal{S}_l, \mathcal{G}\right] = 1$.

Eliminating p_{il} using (65) and (67), we have,

$$2\mathsf{Pr}\left[i \in \bar{G}_{l+1} \mid i \in \mathcal{S}_l, \mathcal{G}\right] + \mathsf{Pr}\left[i \in \bar{G}_l \mid i \in \mathcal{S}_l, \mathcal{G}\right] = 1 \ . \tag{68}$$

Multiplying Eqn. (68) by $\mathsf{Pr}\left[i \in \mathcal{S}_l \mid \mathcal{G}\right]$, we have,

$$2\mathsf{Pr}\left[i \in \bar{G}_{l+1}, i \in \mathcal{S}_l \mid \mathcal{G}\right] + \mathsf{Pr}\left[i \in \bar{G}_l, i \in \mathcal{S}_l \mid \mathcal{G}\right] = \mathsf{Pr}\left[i \in \mathcal{S}_l \mid \mathcal{G}\right] \ . \tag{69}$$

By Lemma 45, if $i \in \text{lmargin}(G_l)$, then, $l_d(i) \not< l$ and $l_d(i) = l$ (or, $|\hat{f}_{il}| \geq Q_l$) iff $i \in \mathcal{S}_l$. By construction therefore, ($i \in \bar{G}_l$ or $i \in \bar{G}_{l+1}$) iff $i \in \mathcal{S}_l$. This proves part 2(b).

Thus, $i \in \bar{G}_{l+1}$ implies $i \in \mathcal{S}_l$ and $i \in \bar{G}_l$ also implies that $i \in \bar{G}_l$. Hence, Eqn. (69) can be written as

$$2\mathsf{Pr}\left[i \in \bar{G}_{l+1} \mid \mathcal{G}\right] + \mathsf{Pr}\left[i \in \bar{G}_l \mid \mathcal{G}\right] = \mathsf{Pr}\left[i \in \mathcal{S}_l \mid \mathcal{G}\right] = 2^{-l} \pm n^{-c} \tag{70}$$

using Fact (43). Multiplying by 2^l gives part 2(c) of the lemma.

Case: $i \in \textit{rmargin}(G_l)$. Assume that $\mathcal{G}$ holds. By Lemma 45, $l_d(i) \in \{l-1, l\}$ but $l_d(i) \not< l-1$ and $l_d(i) \not> l+1$. Since, $l_d(i) \not< l-1$, it follows that $i \notin \bar{G}_r$ for any $r < l-1$. If $i \in \mathcal{S}_l$, we have,

$$|\hat{f}_{il}| \geq |f_i| - \bar{\epsilon}T_l \geq T_{l-1} - 2\bar{\epsilon}T_{l-1} - \bar{\epsilon}T_l = T_l\left((2\alpha)^{1/2} - \bar{\epsilon}(2(2\alpha)^{1/2}+1)\right) \geq (1.3)T_l > T_l$$

by the choice of parameters α and $\bar{\epsilon} = 1/(27p)$. Hence, if $i \notin \bar{G}_{l-1}$ and $i \in \mathcal{S}_l$, then, $i \in \bar{G}_l$. In other words,

$$\mathsf{Pr}\left[i \in \bar{G}_l \mid i \notin \bar{G}_{l-1}, i \in \mathcal{S}_l, \mathcal{G}\right] = 1 \ .$$

If $i \in \bar{G}_r$ for some $r \geq l+1$, then, $i \in \mathcal{S}_l$ and this implies that $i \in \bar{G}_l$, which is a contradiction. Hence,

$$\mathsf{Pr}\left[i \in \cup_{r \notin \{l-1,l\}} \bar{G}_r \mid \mathcal{G}\right] = 0 \ .$$

By construction, we have,

$$\mathsf{Pr}\left[i \in \bar{G}_{l-1} \mid i \in \mathcal{S}_{l-1}, \mathcal{G}\right] = \mathsf{Pr}\left[|\hat{f}_{i,l-1}| \geq T_{l-1} \mid i \in \mathcal{S}_{l-1}, \mathcal{G}\right] = p_{i,l-1} \text{ (say)} \tag{71}$$

$$\begin{aligned}\mathsf{Pr}\left[i \in \bar{G}_l \mid i \in \mathcal{S}_{l-1}, \mathcal{G}\right] &= \mathsf{Pr}\left[Q_{l-1} \leq |\hat{f}_{i,l-1}| < T_{l-1} \text{ and } K_i = 1 \mid i \in \mathcal{S}_{l-1}, \mathcal{G}\right] \\ &\quad + \mathsf{Pr}\left[|\hat{f}_{i,l-1}| < Q_{l-1}, i \in \mathcal{S}_l, |\hat{f}_{il}| \geq T_l \mid i \in \mathcal{S}_{l-1}, \mathcal{G}\right] \\ &= A + B \end{aligned} \tag{72}$$

where, we let A and B denote the probability expressions in the first and second terms in the *RHS* respectively of Eqn. (72). Then,

$$\begin{aligned} A &= \mathsf{Pr}\left[Q_{l-1} \leq |\hat{f}_{i,l-1}| < T_{l-1}, K_i = 1 \mid i \in \mathcal{S}_{l-1}, \mathcal{G}\right] \\ &= \mathsf{Pr}\left[K_i = 1 \mid Q_{l-1} \leq |\hat{f}_{i,l-1}| < T_{l-1}, i \in \mathcal{S}_{l-1}, \mathcal{G}\right] \cdot \mathsf{Pr}\left[Q_{l-1} \leq |\hat{f}_{i,l-1}| < T_{l-1} \mid i \in \mathcal{S}_{l-1}, \mathcal{G}\right] \\ &= (1/2)\mathsf{Pr}\left[Q_{l-1} \leq |\hat{f}_{i,l-1}| < T_{l-1} \mid i \in \mathcal{S}_{l-1}, \mathcal{G}\right] \end{aligned} \tag{73}$$

Therefore, for $i \in \text{rmargin}(G_l)$, i could possibly be a member of $\bar{G}_{l-1}$ which can happen only if $i \in \mathcal{S}_{l-1}$. However, if $i \notin \bar{G}_{l-1}$ and $i \in \mathcal{S}_{l-1}$, then i can possibly be a member of $\bar{G}_l$. This can happen in two ways, either (i) $Q_{l-1} \leq |\hat{f}_{i,l-1}| < T_{l-1}$ and the coin toss $K_i = 1$, or, (ii) $Q_{l-1} > |\hat{f}_{i,l-1}|$ and $i \in \mathcal{S}_l$ and $|\hat{f}_{il}| \geq T_l$. In the latter case, if $i \in \mathcal{S}_l$, then, $|\hat{f}_{il}|$ is at least T_l with probability 1, conditional on $\mathcal{G}$. This follows from Lemma 45, part (2). In particular, $i \notin \bar{G}_{l'}$ for any $l' \notin \{l-1, l\}$.

Hence,

$$\begin{aligned} B &= \mathsf{Pr}\left[|\hat{f}_{i,l-1}| < Q_{l-1}, i \in \mathcal{S}_l, |\hat{f}_{il}| \geq T_l \mid i \in \mathcal{S}_{l-1}, \mathcal{G}\right] \\ &= \mathsf{Pr}\left[|\hat{f}_{i,l-1}| < Q_{l-1}, i \in \mathcal{S}_l \mid i \in \mathcal{S}_{l-1}, \mathcal{G}\right] \\ &= \mathsf{Pr}\left[|\hat{f}_{i,l-1}| < Q_{l-1} \mid i \in \mathcal{S}_l, \mathcal{G}\right] \cdot \mathsf{Pr}\left[i \in \mathcal{S}_l \mid i \in \mathcal{S}_{l-1}, \mathcal{G}\right] \\ &= \mathsf{Pr}\left[|\hat{f}_{i,l-1}| < Q_{l-1} \mid i \in \mathcal{S}_l, \mathcal{G}\right] \cdot \left(1/2 \pm n^{-c}\right) \end{aligned}$$

Note that $\mathsf{Pr}\big[|\hat{f}_{i,l-1}| < Q_{l-1} \mid i \in \mathcal{S}_l\big] = \mathsf{Pr}\big[|\hat{f}_{i,l-1}| < Q_{l-1} \mid i \in \mathcal{S}_{l-1}\big]$ for the following reason. $|\hat{f}_{i,l-1}|$ is a function of the frequencies of the items that conflict with i in the set of hash buckets to which i maps in the HH_{l-1} structure. By construction of the hash function, whether i maps to the next level l depends on whether $g_l(i) = 1$, which is independent of the hash functions $g_1, g_2, \ldots, g_{l-1}$. Hence,

$$\mathsf{Pr}\big[|\hat{f}_{i,l-1}| < Q_{l-1} \mid i \in \mathcal{S}_l\big] = \mathsf{Pr}\big[|\hat{f}_{i,l-1}| < Q_{l-1} \mid i \in \mathcal{S}_{l-1}\big]$$

Using Fact (43), we have,

$$\mathsf{Pr}\left[|\hat{f}_{i,l-1}| < Q_{l-1} \mid i \in \mathcal{S}_l, \mathcal{G}\right] = \mathsf{Pr}\left[|\hat{f}_{i,l-1}| < Q_{l-1} \mid i \in \mathcal{S}_{l-1}, \mathcal{G}\right] \pm n^{-c} \ .$$

Thus Eqn. (72) may be written as

$$\begin{aligned} &\mathsf{Pr}\big[i \in \bar{G}_l \mid i \in \mathcal{S}_{l-1}, \mathcal{G}\big] = A + B \\ &= (1/2)\mathsf{Pr}\big[Q_{l-1} \leq |\hat{f}_{i,l-1}| < T_{l-1} \mid i \in \mathcal{S}_{l-1}, \mathcal{G}\big] + (1/2)\mathsf{Pr}\big[|\hat{f}_{i,l-1}| < Q_{l-1} \mid i \in \mathcal{S}_{l-1}, \mathcal{G}\big] \pm O(n^{-c}) \\ &= (1/2)\mathsf{Pr}\big[|\hat{f}_{i,l-1}| < T_{l-1} \mid i \in \mathcal{S}_{l-1}, \mathcal{G}\big] \pm O(n^{-c}) \\ &= \frac{1 - p_{i,l-1}}{2} \pm O(n^{-c}) \ . \end{aligned} \tag{74}$$

From Eqns. (71) and (74) we obtain,

$$2\mathsf{Pr}\left[i \in \bar{G}_l \mid i \in \mathcal{S}_{l-1}, \mathcal{G}\right] + \mathsf{Pr}\left[i \in \bar{G}_{l-1} \mid i \in \mathcal{S}_{l-1}, \mathcal{G}\right] = 1 \pm O(n^{-c}) \ . \tag{75}$$

Multiplying Eqn. (75) by $\mathsf{Pr}\left[i \in \mathcal{S}_{l-1} \mid \mathcal{G}\right]$, we have,

$$2\mathsf{Pr}\left[i \in \bar{G}_l, i \in \mathcal{S}_{l-1} \mid \mathcal{G}\right] + \mathsf{Pr}\left[i \in \bar{G}_{l-1}, i \in \mathcal{S}_{l-1} \mid \mathcal{G}\right] = \mathsf{Pr}\left[i \in \mathcal{S}_{l-1} \mid \mathcal{G}\right]\left(1 \pm O(n^{-c})\right) \ . \tag{76}$$

From the discussion after Eqn. (73), it follows that i may belong to $\bar{G}_{l-1} \cup \bar{G}_l$, and in either case, this is possible only if $i \in \mathcal{S}_{l-1}$. This proves part 3 (b).

Thus, $i \in \bar{G}_l$ or $i \in \bar{G}_{l-1}$ implies that $i \in \mathcal{S}_{l-1}$. Hence, Eqn. (76) is equivalent to

$$\begin{aligned} 2\mathsf{Pr}\left[i \in \bar{G}_l \mid \mathcal{G}\right] + \mathsf{Pr}\left[i \in \bar{G}_{l-1} \mid \mathcal{G}\right] &= (2^{-(l-1)} \pm n^{-c})\left(1 \pm O(n^{-c})\right) \\ &= 2^{-(l-1)} \pm O(n^{-c}) \end{aligned}$$

Multiplying by 2^{l-1} gives statement 3 (c) of the lemma.

□

E Approximate pair-wise independence of the sampling

In this section, we prove an approximate pair-wise independence property of the sampling technique.

Lemma 46. *Let $i \neq j$. Then, $\mathsf{Pr}\left[i \in \mathcal{S}_l \mid j \in \mathcal{S}_r, \mathcal{G}\right] = 2^{-l} \pm n^{-c}$.*

Proof. By pair-wise independence of the hash functions $\{g_l\}$ mapping items to levels, we have $\mathsf{Pr}\left[i \in \mathcal{S}_l \mid j \in \mathcal{S}_r\right] = \mathsf{Pr}\left[i \in \mathcal{S}_l\right] = 2^{-l}$. By Fact 43, $\mathsf{Pr}\left[i \in \mathcal{S}_l \mid j \in \mathcal{S}_r, \mathcal{G}\right] = 2^{-l} \pm n^{-c}$. □

E.1 Sampling probability of items conditional on another item mapping to a level

Restated Lemma (Restatement of Lemma 10.)**.** *Let $i, j \in [n]$, $i \neq j$ and $j \in G_r$. Then,*

$$\sum_{r'=0}^{L} 2^{r'} \mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_l, \mathcal{G}\right] = 1 \pm O(2^r \cdot n^{-c}) \ .$$

In particular, the following hold.

1) Suppose $j \in mid(G_r)$. Then,

$$2^r \mathsf{Pr}\left[j \in \bar{G}_r \mid i \in \mathcal{S}_l, \mathcal{G}\right] = 1 \pm 2^r n^{-c} \ .$$

Further, for any $r \neq r'$, $\mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_l, \mathcal{G}\right] = 0$.

2) If $j \in lmargin(G_r)$, then,

$$2^{r+1} \mathsf{Pr}\left[j \in \bar{G}_{r+1} \mid i \in \mathcal{S}_l, \mathcal{G}\right] + 2^r \mathsf{Pr}\left[j \in \bar{G}_r \mid i \in \mathcal{S}_l, \mathcal{G}\right] = 1 \pm 2^{r+1} n^{-c} \ .$$

Further, for any $r' \notin \{r, r+1\}$, $\mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_l, \mathcal{G}\right] = 0$.

3) If $j \in rmargin(G_r)$, then

$$2^r \mathsf{Pr}\left[j \in \bar{G}_r \mid i \in \mathcal{S}_l, \mathcal{G}\right] + 2^{r-1} \mathsf{Pr}\left[j \in \bar{G}_{r-1} \mid i \in \mathcal{S}_l, \mathcal{G}\right] = 1 \pm 2^{r+1} n^{-c} \ .$$

Further, for any $r' \notin \{r-1, r\}$, $\mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_l, \mathcal{G}\right] = 0$.

Proof of Lemma 10. The proof proceeds identically as in the proof of Lemma 8, except that all probabilities are, in addition to being conditional on $\mathcal{G}$, also conditional on $i \in \mathcal{S}_l$.

Case 1: $j \in mid(G_r)$. Conditional on $\mathcal{G}$, as argued in the proof of Lemma 45, part 1 (b), $j \in \bar{G}_r$ iff $j \in \mathcal{S}_r$. Therefore,

$$\mathsf{Pr}\left[j \in \bar{G}_r \mid i \in \mathcal{S}_l, \mathcal{G}\right] = \mathsf{Pr}\left[j \in \mathcal{S}_r \mid i \in \mathcal{S}_l, \mathcal{G}\right] \in 2^{-r} \pm n^{-c} \tag{77}$$

where, the last step follows from Lemma 46.

Case 2: $j \in lmargin(G_r)$. Let

$$p'_{jr} = \mathsf{Pr}\left[|\hat{f}_{ir}| \geq T_r \mid i \in \mathcal{S}_l, j \in \mathcal{S}_r, \mathcal{G}\right] \ .$$

Then,

$$\begin{aligned}\mathsf{Pr}\big[j \in \bar{G}_r \mid i \in \mathcal{S}_l, \mathcal{G}\big] &= \mathsf{Pr}\left[|\hat{f}_{ir}| \geq T_r, j \in \mathcal{S}_r \mid i \in \mathcal{S}_l, \mathcal{G}\right] \\ &= \mathsf{Pr}\left[|\hat{f}_{ir}| \geq T_r \mid i \in \mathcal{S}_l, j \in \mathcal{S}_r, \mathcal{G}\right] \mathsf{Pr}\left[j \in \mathcal{S}_r \mid i \in \mathcal{S}_l, \mathcal{G}\right] \\ &= p'_{jr} \cdot (2^{-r} \pm n^{-c}), \quad \text{by Lemma 46.} \end{aligned} \tag{78}$$

Further,

$$\begin{aligned}\mathsf{Pr}\big[j \in \bar{G}_{r+1} \mid i \in \mathcal{S}_l, \mathcal{G}\big] &= \mathsf{Pr}\left[Q_r \leq |\hat{f}_{ir}| < T_r, j \in \mathcal{S}_r, K_i = 1 \mid i \in \mathcal{S}_l, \mathcal{G}\right] \\ &\quad + \mathsf{Pr}\left[|\hat{f}_{ir}| < Q_r, i \in \mathcal{S}_{r+1}, |\hat{f}_{i,r+1}| \geq T_{r+1} \mid i \in \mathcal{S}_l, \mathcal{G}\right] \end{aligned} \tag{79}$$

Conditional on $\mathcal{G}$, $|\hat{f}_{ir}| \geq |f_i| - \bar{\epsilon}T_r \geq T_r - \bar{\epsilon}T_r = Q_r$, since $j \in \text{lmargin}(G_r)$. Hence, $\mathsf{Pr}\left[|\hat{f}_{ir}| < Q_r \mid \mathcal{G}\right] = 0$. Further, since the coin toss $K_i = 1$ is independent of other random bits, Eqn. (79) becomes

$$\begin{aligned}\mathsf{Pr}\left[j \in \bar{G}_{r+1} \mid i \in \mathcal{S}_l, \mathcal{G}\right] &= (1/2)\mathsf{Pr}\left[Q_r \leq |\hat{f}_{ir}| < T_r, j \in \mathcal{S}_r \mid i \in \mathcal{S}_l, \mathcal{G}\right] \\ &= (1/2)\mathsf{Pr}\left[Q_r \leq |\hat{f}_{ir}| < T_r \mid i \in \mathcal{S}_l, j \in \mathcal{S}_r, \mathcal{G}\right] \mathsf{Pr}\left[j \in \mathcal{S}_r \mid i \in \mathcal{S}_l, \mathcal{G}\right] \\ &= (1/2)(1 - p'_{jr})(2^{-r} \pm n^{-c}) \end{aligned} \tag{80}$$

Multiplying Eqn. (80) by 2^{r+1}, multiplying Eqn. (79) by 2^r and adding, we have,

$$2^{r+1}\mathsf{Pr}\left[j \in \bar{G}_{r+1} \mid i \in \mathcal{S}_l, \mathcal{G}\right] + 2^r \mathsf{Pr}\left[j \in \bar{G}_r \mid i \in \mathcal{S}_l, \mathcal{G}\right] = 1 \pm O(2^r n^{-c})$$

which proves statement (2) of the lemma.

Case 3: $j \in \textit{rmargin}(G_r)$. Then,

$$\begin{aligned}\mathsf{Pr}\left[j \in \bar{G}_{r-1} \mid i \in \mathcal{S}_l, \mathcal{G}\right] &= \mathsf{Pr}\left[|\hat{f}_{j,r-1}| \geq T_{r-1}, j \in \mathcal{S}_{r-1}, \mid i \in \mathcal{S}_l, \mathcal{G}\right] \\ &= \mathsf{Pr}\left[|\hat{f}_{j,r-1}| \geq T_{r-1} \mid i \in \mathcal{S}_l, j \in \mathcal{S}_{r-1}, \mathcal{G}\right] \cdot \mathsf{Pr}\left[j \in \mathcal{S}_{r-1} \mid i \in \mathcal{S}_l, \mathcal{G}\right] \\ &= \mathsf{Pr}\left[|\hat{f}_{j,r-1}| \geq T_{r-1} \mid i \in \mathcal{S}_l, j \in \mathcal{S}_{r-1}, \mathcal{G}\right] (2^{-(r-1)} \pm n^{-c}) \end{aligned} \tag{81}$$

Also,

$$\begin{aligned}\mathsf{Pr}\left[j \in \bar{G}_r \mid i \in \mathcal{S}_l, \mathcal{G}\right] &= \mathsf{Pr}\left[j \in \mathcal{S}_{r-1}, Q_{r-1} \leq |\hat{f}_{j,r-1}| < T_{r-1}, K_i = 1 \mid i \in \mathcal{S}_l, \mathcal{G}\right] \\ &\quad + \mathsf{Pr}\left[|\hat{f}_{j,r-1}| < Q_r, j \in \mathcal{S}_r, |\hat{f}_{j,r}| \geq T_r \mid i \in \mathcal{S}_l, \mathcal{G}\right] \end{aligned} \tag{82}$$

For $j \in \text{rmargin}(G_r)$ and conditional on $\mathcal{G}$, by following the argument of Lemma 45, it follows that if $j \in \mathcal{S}_r$ then, $|\hat{f}_{jr}| \geq T_r$, viz., $|\hat{f}_{jr}| \geq |f_{jr}| - \bar{\epsilon}T_r \geq T_{r-1} - 2\bar{\epsilon}T_{r-1} - \bar{\epsilon}T_r > T_r$. Therefore,

$$\begin{aligned}&\mathsf{Pr}\left[|\hat{f}_{j,r-1}| < Q_r, j \in \mathcal{S}_r, |\hat{f}_{j,r}| \geq T_r \mid i \in \mathcal{S}_l, \mathcal{G}\right] \\ &= \mathsf{Pr}\left[|\hat{f}_{j,r-1}| < Q_r, j \in \mathcal{S}_r \mid i \in \mathcal{S}_l, \mathcal{G}\right] \\ &= \mathsf{Pr}\left[|\hat{f}_{j,r-1}| < Q_r \mid i \in \mathcal{S}_l, j \in \mathcal{S}_r, \mathcal{G}\right] \mathsf{Pr}\left[j \in \mathcal{S}_r \mid i \in \mathcal{S}_l, \mathcal{G}\right] \\ &= \mathsf{Pr}\left[|\hat{f}_{j,r-1}| < Q_r \mid i \in \mathcal{S}_l, j \in \mathcal{S}_r, \mathcal{G}\right] (2^{-r} \pm n^{-c}) \end{aligned} \tag{83}$$

The estimate $\hat{f}_{j,r-1}$ is obtained at level $r-1$, and this is independent of whether j (or any other subset of items) is a member of $\mathcal{S}_r$. The latter is a consequence of the level-wise product of *independent* hash values, namely, $j \in \mathcal{S}_r$ iff $j \in \mathcal{S}_{r-1}$ and $g_r(j) = 1$. Therefore,

$$\begin{aligned}
&\mathsf{Pr}\left[|\hat{f}_{j,r-1}| < Q_r \mid i \in \mathcal{S}_l, j \in \mathcal{S}_r\right] \\
&= \mathsf{Pr}\left[|\hat{f}_{j,r-1}| < Q_r \mid i \in \mathcal{S}_l, j \in \mathcal{S}_{r-1}, g_r(j) = 1\right] \\
&= \frac{\mathsf{Pr}\left[|\hat{f}_{j,r-1}| < Q_r, g_r(j) = 1 \mid i \in \mathcal{S}_l, j \in \mathcal{S}_{r-1}\right]}{\mathsf{Pr}\left[g_r(j) = 1 \mid i \in \mathcal{S}_l, j \in \mathcal{S}_{r-1}\right]} \\
&= \left(\frac{\mathsf{Pr}\left[g_r(j) = 1 \mid |\hat{f}_{j,r-1}| < Q_r, i \in \mathcal{S}_l, j \in \mathcal{S}_{r-1}\right]}{\mathsf{Pr}\left[g_r(j) = 1 \mid i \in \mathcal{S}_l, j \in \mathcal{S}_{r-1}\right]}\right) \cdot \left(\mathsf{Pr}\left[|\hat{f}_{j,r-1}| < Q_r \mid i \in \mathcal{S}_l, j \in \mathcal{S}_{r-1}\right]\right) \quad (84)
\end{aligned}$$

Consider the numerator term of the fraction above:
$\mathsf{Pr}\left[g_r(j) = 1 \mid |\hat{f}_{j,r-1}| < Q_r, i \in \mathcal{S}_l, j \in \mathcal{S}_{r-1}\right]$. The event $|\hat{f}_{j,r-1}| < Q_r$ depends only on the set of elements that have mapped to $\mathcal{S}_{r-1}$, and is independent of whether $g_r(j) = 1$. Similarly, $j \in \mathcal{S}_{r-1}$ is independent of whether $g_r(j) = 1$. Thus, the numerator term equals $\mathsf{Pr}\left[g_r(j) = 1 \mid i \in \mathcal{S}_l\right]$ and the denominator term also equals the same, for the same reasons. Hence, Eqn. (84) becomes

$$\mathsf{Pr}\left[|\hat{f}_{j,r-1}| < Q_r \mid i \in \mathcal{S}_l, j \in \mathcal{S}_r\right] = \mathsf{Pr}\left[|\hat{f}_{j,r-1}| < Q_r \mid i \in \mathcal{S}_l, j \in \mathcal{S}_{r-1}\right] \quad (85)$$

Now, conditioning with respect to $\mathcal{G}$, we have,

$$\mathsf{Pr}\left[\hat{f}_{j,r-1} > Q_r \mid j \in \mathcal{S}_r, i \in \mathcal{S}_l, \mathcal{G}\right] \in \mathsf{Pr}\left[\hat{f}_{i,r-1} > Q_r \mid j \in \mathcal{S}_{r-1}, i \in \mathcal{S}_l, \mathcal{G}\right] \pm n^{-c} \ . \quad (86)$$

Substituting Eqn. (86) in Eqn. (83), we have,

$$\begin{aligned}
&\mathsf{Pr}\left[|\hat{f}_{j,r-1}| < Q_r, j \in \mathcal{S}_r, |\hat{f}_{j,r}| \geq T_r \mid i \in \mathcal{S}_l, \mathcal{G}\right] \\
&= \left(\mathsf{Pr}\left[\hat{f}_{j,r-1} < Q_r \mid j \in \mathcal{S}_{r-1}, i \in \mathcal{S}_l, \mathcal{G}\right]\right)(2^{-r} \pm n^{-c}) \pm 2^{-r}n^{-c} \quad (87)
\end{aligned}$$

Consider the first probability term in the *RHS* of Eqn. (82).

$$\begin{aligned}
&\mathsf{Pr}\left[j \in \mathcal{S}_{r-1}, Q_{r-1} \leq |\hat{f}_{j,r-1}| < T_{r-1}, K_i = 1 \mid i \in \mathcal{S}_l, \mathcal{G}\right] \\
&= (1/2)\mathsf{Pr}\left[Q_{r-1} \leq |\hat{f}_{j,r-1}| < T_{r-1} \mid i \in \mathcal{S}_l, j \in \mathcal{S}_{r-1}, \mathcal{G}\right] \mathsf{Pr}\left[j \in \mathcal{S}_{r-1} \mid i \in \mathcal{S}_l, \mathcal{G}\right] \\
&= \mathsf{Pr}\left[Q_{r-1} \leq |\hat{f}_{j,r-1}| < T_{r-1} \mid i \in \mathcal{S}_l, j \in \mathcal{S}_{r-1}, \mathcal{G}\right](1/2)(2^{-(r-1)} \pm n^{-c}) \quad (88)
\end{aligned}$$

Substituting Eqns. (87) and (88) in Eqn. (82), we have,

$$\begin{aligned}
&\mathsf{Pr}\left[j \in \bar{G}_r \mid i \in \mathcal{S}_l, \mathcal{G}\right] \\
&\quad = \mathsf{Pr}\left[Q_{r-1} \leq |\hat{f}_{j,r-1}| < T_{r-1} \mid i \in \mathcal{S}_l, j \in \mathcal{S}_{r-1}, \mathcal{G}\right] 2^{-r} \pm O(n^{-c}) \\
&\qquad + \left(\mathsf{Pr}\left[\hat{f}_{i,r-1} < Q_r \mid j \in \mathcal{S}_{r-1}, i \in \mathcal{S}_l, \mathcal{G}\right]\right)(2^{-r} \pm n^{-c}) \pm 2^{-r}n^{-c} \quad (89)
\end{aligned}$$

Multiplying Eqn. (81) by 2^{r-1} and Eqn. (89) by 2^r and adding, we obtain

$$\begin{aligned}
&2^{r-1}\mathsf{Pr}\left[j \in \bar{G}_{r-1} \mid i \in \mathcal{S}_l, \mathcal{G}\right] + 2^r \mathsf{Pr}\left[j \in \bar{G}_r \mid i \in \mathcal{S}_l, \mathcal{G}\right] \\
&= \mathsf{Pr}\left[|\hat{f}_{j,r-1}| \geq T_{l-1} \mid i \in \mathcal{S}_l, j \in \mathcal{S}_{r-1}, \mathcal{G}\right] \pm 2^{r-1} n^{-c} \\
&\quad + \mathsf{Pr}\left[Q_{r-1} \leq |\hat{f}_{j,r-1}| < T_{r-1} \mid i \in \mathcal{S}_l, j \in \mathcal{S}_{r-1}, \mathcal{G}\right] \pm O(2^r n^{-c})) \\
&\quad\quad + \mathsf{Pr}\left[\hat{f}_{i,r-1} < Q_r \mid j \in \mathcal{S}_{r-1}, i \in \mathcal{S}_l, \mathcal{G}\right] \pm O(2^r n^{-c}) \\
&= 1 \pm O(2^r n^{-c}) \ .
\end{aligned}$$

This proves statement (3) of the Lemma. □

E.2 Sampling probability of an item conditional on another item being sampled

Restated Lemma (Lemma 12.)**.** *Suppose $i \in G_l$, $j \in G_m$ and $j \neq i$. Then,*

$$\sum_{r,r'=0}^{L} 2^{r+r'} \mathsf{Pr}\left[i \in \bar{G}_r, j \in \bar{G}_{r'} \mid \mathcal{G}\right] = 1 \pm O((2^l + 2^m) n^{-c}) \ .$$

Proof of Lemma 12. Assume $\mathcal{G}$ holds for all the arguments in the proof. *Case 1: $i \in mid(G_l)$.* Then,

$$\mathsf{Pr}\left[i \in \bar{G}_r, j \in \bar{G}_{r'} \mid \mathcal{G}\right] = \mathsf{Pr}\left[i \in \bar{G}_r \mid j \in \bar{G}_{r'}, \mathcal{G}\right] \cdot \mathsf{Pr}\left[j \in \bar{G}_{r'} \mid \mathcal{G}\right]$$

Conditional on $\mathcal{G}$, $i \in \bar{G}_r$ iff $r = l$ and $i \in \mathcal{S}_l$. That is, for $r \neq l$, $\mathsf{Pr}\left[i \in \bar{G}_r \mid j \in \bar{G}_{r'}, \mathcal{G}\right] = 0$. Therefore,

$$\begin{aligned}
&\mathsf{Pr}\left[i \in \bar{G}_l \mid j \in \bar{G}_{r'}, \mathcal{G}\right] \cdot \mathsf{Pr}\left[j \in \bar{G}_{r'} \mid \mathcal{G}\right] \\
&= \mathsf{Pr}\left[i \in \mathcal{S}_l \mid j \in \bar{G}_r, \mathcal{G}\right] \cdot \mathsf{Pr}\left[j \in \bar{G}_{r'} \mid \mathcal{G}\right], \text{ by Lemma 45, part 2(b)} \\
&= \mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_l, \mathcal{G}\right] \cdot \mathsf{Pr}[i \in \mathcal{S}_l \mid \mathcal{G}] \quad \text{, by Bayes' rule} \\
&= \mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_l, \mathcal{G}\right] \cdot (2^{-l} \pm n^{-c})
\end{aligned}$$

Multiplying by 2^l, we have,

$$2^l \mathsf{Pr}\left[i \in \bar{G}_r, j \in \bar{G}_{r'} \mid \mathcal{G}\right] = \mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_l, \mathcal{G}\right] (1 \pm 2^l n^{-c}) \tag{90}$$

By Lemma 10, we have, $\sum_{r'=0}^{L} \mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_l, \mathcal{G}\right] = 1 \pm 2^{m+1} n^{-c}$. Therefore, multiplying both sides of Eqn. (90) by $2^{r'}$ and summing over r', we have,

$$\sum_{r'=0}^{L} 2^{l+r'} \mathsf{Pr}\left[i \in \bar{G}_l, j \in \bar{G}_{r'} \mid \mathcal{G}\right] = (1 \pm 2^{m+1} n^{-c})(1 \pm 2^l n^{-c}) = (1 \pm O(2^m + 2^l) n^{-c}) \tag{91}$$

Since $\mathsf{Pr}\left[i \in \bar{G}_r, j \in \bar{G}_{r'} \mid \mathcal{G}\right] = 0$ for $r \neq l$, we can equivalently write Eqn. (91) as

$$\sum_{r,r'=0}^{L} 2^{r+r'} \mathsf{Pr}\left[i \in \bar{G}_r, j \in \bar{G}_{r'} \mid \mathcal{G}\right] = (1 \pm O(2^m + 2^l) n^{-c})$$

Case 2: $i \in lmargin(G_l)$. Then, i may belong to either $\bar{G}_l \cup \bar{G}_{l+1}$ and to no other sampled group and $i \in \bar{G}_l \cup \bar{G}_{l+1}$ iff $i \in \mathcal{S}_l$, by Lemma 8 parts 2(a) and 2(b) respectively.

$$\begin{aligned}
&\mathsf{Pr}\left[i \in \bar{G}_l, j \in \bar{G}_{r'} \mid \mathcal{G}\right] \\
&= \mathsf{Pr}\left[i \in \mathcal{S}_l, |\hat{f}_{il}| \geq T_l, j \in \bar{G}_{r'} \mid \mathcal{G}\right] \mathsf{Pr}\left[j \in \bar{G}_{r'}, \mathcal{G}\right] \\
&= \mathsf{Pr}\left[|\hat{f}_{il}| \geq T_l \mid j \in \bar{G}_{r'}, i \in \mathcal{S}_l, \mathcal{G}\right] \mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_l, \mathcal{G}\right] \mathsf{Pr}\left[i \in \mathcal{S}_l \mid \mathcal{G}\right] \\
&= \mathsf{Pr}\left[|\hat{f}_{il}| \geq T_l \mid j \in \bar{G}_{r'}, i \in \mathcal{S}_l, \mathcal{G}\right] \mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_l, \mathcal{G}\right] (2^{-l} \pm n^{-c}) \qquad (92)
\end{aligned}$$

Let

$$p_{il} = \mathsf{Pr}\left[|\hat{f}_{il}| \geq T_l \mid j \in \bar{G}_{r'}, i \in \mathcal{S}_l, \mathcal{G}\right] \ .$$

Multiplying both sides of Eqn. (92) by 2^l, we obtain

$$2^l \mathsf{Pr}\left[i \in \bar{G}_l, j \in \bar{G}_{r'} \mid \mathcal{G}\right] = p_{il} \cdot \mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_l, \mathcal{G}\right] (1 \pm 2^l n^{-c}) \qquad (93)$$

We now consider the case when $i \in \bar{G}_{l+1}$. By construction, $i \in \bar{G}_{l+1}$ in two ways, either (i) $i \in \mathcal{S}_l$, $Q_l \leq |\hat{f}_{il}| < T_l$ and $K_i = 1$, or, (ii) $i \in \mathcal{S}_l$, $|\hat{f}_{il}| < Q_l$ and $i \in \mathcal{S}_l$ and $|\hat{f}_{i,l+1}| \geq T_{l+1}$. Possibility (ii) cannot hold since, by Lemma 45 (1b), $i \in \mathcal{S}_l$ iff $l_d(i) = l$, which by definition is that $|\hat{f}_{il}| \geq Q_l$. These calculations are conditioned on $\mathcal{G}$ and therefore hold conditioned on $j \in \bar{G}_{r'}$ as well. Hence,

$$\begin{aligned}
&\mathsf{Pr}\left[i \in \bar{G}_{l+1}, j \in \bar{G}_{r'} \mid \mathcal{G}\right] \\
&= \mathsf{Pr}\left[i \in \mathcal{S}_l, (Q_l \leq |\hat{f}_{il}| < T_l), K_i = 1, j \in \bar{G}_{r'} \mid \mathcal{G}\right] \\
&= (1/2)\mathsf{Pr}\left[Q_l \leq |\hat{f}_{il}| < T_l \mid i \in \mathcal{S}_l, j \in \bar{G}_{r'}, \mathcal{G}\right] \mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_l, \mathcal{G}\right] \mathsf{Pr}\left[i \in \mathcal{S}_l \mid \mathcal{G}\right] \\
&= (1/2)(1 - p_{il})\mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_l, \mathcal{G}\right] (2^{-l} \pm n^{-c}) \qquad (94)
\end{aligned}$$

or, by multiplying both sides of Eqn. (94),

$$2^{l+1}\mathsf{Pr}\left[i \in \bar{G}_{l+1}, j \in \bar{G}_{r'} \mid \mathcal{G}\right] = (1 - p_{il}) \cdot \mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_l, \mathcal{G}\right] (1 \pm 2^{l+1} n^{-c}) \qquad (95)$$

Adding Eqns. (93) and (95), we have,

$$2^{l+1}\mathsf{Pr}\left[i \in \bar{G}_{l+1}, j \in \bar{G}_{r'} \mid \mathcal{G}\right] + 2^l \mathsf{Pr}\left[i \in \bar{G}_l, j \in \bar{G}_{r'} \mid \mathcal{G}\right] = \mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_l, \mathcal{G}\right] (1 \pm 2^{l+2} n^{-c}) \ . \qquad (96)$$

By Lemma 10, $\sum_{r'=0}^{L} 2^{r'} \mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_l, \mathcal{G}\right] = 1 \pm O(2^m n^{-c})$. Therefore, multiplying Eqn. (96) by $2^{r'}$ and summing over r', we have,

$$\begin{aligned}
&\sum_{r'=0}^{L} \left(2^{l+1}\mathsf{Pr}\left[i \in \bar{G}_{l+1}, j \in \bar{G}_{r'} \mid \mathcal{G}\right] + 2^l \mathsf{Pr}\left[i \in \bar{G}_l, j \in \bar{G}_{r'} \mid \mathcal{G}\right]\right) \\
&= \sum_{r'=0}^{L} 2^{r'} \mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_l, \mathcal{G}\right] (1 \pm 2^{l+2} n^{-c}) \\
&= (1 \pm O(2^m n^{-c}))(1 \pm O(2^l n^{-c})) \\
&= 1 \pm O((2^l + 2^m) n^{-c})
\end{aligned}$$

Since, $\mathsf{Pr}\left[i \in \bar{G}_r, j \in \bar{G}_{r'} \mid \mathcal{G}\right] = 0$ for any $r \notin \{l, l+1\}$, we can rewrite the above equation as

$$\sum_{r,r'=0}^{L} 2^{r+r'} \mathsf{Pr}\left[i \in \bar{G}_r, j \in \bar{G}_{r'} \mid \mathcal{G}\right] = (1 \pm O(2^m + 2^l)n^{-c})$$

Case 3: $i \in rmargin(G_l)$. If $j \in \text{lmargin}(G_m)$ or $j \in \text{mid}(G_m)$, then, we can interchange the roles of i and j and the lemma is proved. Hence, we may now assume that $j \in \text{rmargin}(G_m)$. Let $m \leq l$ without loss of generality.

By Lemma 8, part (3), $i \in \bar{G}_{l-1} \cup \bar{G}_l$ and this implies that $i \in \mathcal{S}_l$. Also, $i \notin \cup_{l' \notin \{l-1,l\}} \bar{G}_{l'}$ (with prob. 1). Let $p_{i,l-1,j,r'} = \mathsf{Pr}\left[|\hat{f}_{i,l-1}| \geq T_{l-1} \mid j \in \bar{G}_{r'}, i \in \mathcal{S}_{l-1}\right]$. Then,

$$\begin{aligned}
\mathsf{Pr}\left[i \in \bar{G}_{l-1}, j \in \bar{G}_{r'} \mid \mathcal{G}\right] &= \mathsf{Pr}\left[|\hat{f}_{i,l-1}| \geq T_{l-1}, i \in \mathcal{S}_{l-1}, j \in \bar{G}_{r'} \mid \mathcal{G}\right] \\
&= p_{i,l-1} \cdot \mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_{l-1}, \mathcal{G}\right] \mathsf{Pr}\left[i \in \mathcal{S}_{l-1} \mid \mathcal{G}\right] \\
&= p_{i,l-1} \cdot \mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_{l-1}, \mathcal{G}\right] (2^{-(l-1)} \pm n^{-c}) \ . \qquad (97)
\end{aligned}$$

Let $q_{i,l-1,j,r'} = \mathsf{Pr}\left[Q_{l-1} \leq |\hat{f}_{i,l-1}| < T_{l-1} \mid i \in \mathcal{S}_{l-1}, j \in \bar{G}_{r'}\right]$. By Lemma 8 part 3 (b), $\{i \in \mathcal{S}_l, l_d(i) \neq l-1\} \subset \{i \in \bar{G}_l\}$. Then,

$$\begin{aligned}
&\mathsf{Pr}\left[i \in \bar{G}_l, j \in \bar{G}_{r'} \mid \mathcal{G}\right] \\
&= \mathsf{Pr}\left[Q_{l-1} \leq |\hat{f}_{i,l-1}| < T_{l-1}, K_i = 1, i \in \mathcal{S}_{l-1}, j \in \bar{G}_{r'} \mid \mathcal{G}\right] + \mathsf{Pr}\left[|\hat{f}_{i,l-1}| < Q_{l-1}, i \in \mathcal{S}_l, j \in \bar{G}_{r'} \mid \mathcal{G}\right] \\
&= (1/2) q_{i,l-1,j,r'} \mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_{l-1}, \mathcal{G}\right] \mathsf{Pr}\left[i \in \mathcal{S}_{l-1}\right] + \mathsf{Pr}\left[|\hat{f}_{i,l-1}| < Q_{l-1}, i \in \mathcal{S}_l, j \in \bar{G}_{r'} \mid \mathcal{G}\right] \\
&= q_{i,l-1,j,r'} \cdot \mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_{l-1}, \mathcal{G}\right] (2^{-l} \pm O(n^{-c}) \\
&\qquad + \mathsf{Pr}\left[|\hat{f}_{i,l-1}| < Q_{l-1} \mid i \in \mathcal{S}_l, j \in \bar{G}_{r'}, \mathcal{G}\right] \mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_l, \mathcal{G}\right] \mathsf{Pr}\left[i \in \mathcal{S}_l \mid \mathcal{G}\right] \qquad (98)
\end{aligned}$$

Consider the following term derived from the second term in the above sum.

$$\begin{aligned}
\mathsf{Pr}\left[|\hat{f}_{i,l-1}| < Q_{l-1} \mid i \in \mathcal{S}_l, j \in \bar{G}_{r'}, \mathcal{G}\right] &= \mathsf{Pr}\left[|\hat{f}_{i,l-1}| < Q_{l-1} \mid g_l(i) = 1, i \in \mathcal{S}_{l-1}, j \in \bar{G}_{r'}\right] \\
&= \frac{\mathsf{Pr}\left[|\hat{f}_{i,l-1}| < Q_{l-1}, g_l(i) = 1 \mid i \in \mathcal{S}_{l-1}, j \in \bar{G}_{r'}\right]}{\mathsf{Pr}\left[g_l(i) = 1 \mid i \in \mathcal{S}_{l-1}, j \in \bar{G}_{r'}\right]} \qquad (99)
\end{aligned}$$

$$\begin{aligned}
&\mathsf{Pr}\left[|\hat{f}_{i,l-1}| < Q_{l-1}, g_l(i) = 1 \mid i \in \mathcal{S}_{l-1}, j \in \bar{G}_{r'}, \mathcal{G}\right] \\
&= \mathsf{Pr}\left[g_l(i) = 1 \mid i \in \mathcal{S}_{l-1}, |\hat{f}_{i,l-1}| < Q_{l-1}, j \in \bar{G}_{r'}, \mathcal{G}\right] \mathsf{Pr}\left[|\hat{f}_{i,l-1}| < Q_{l-1} \mid i \in \mathcal{S}_{l-1}, j \in \bar{G}_{r'}, \mathcal{G}\right]
\end{aligned}$$

The event $g_l(i) = 1$ is independent of the value of $\hat{f}_{i,l-1}$, since they depend on the values of $g_{l'}(k)$'s for $k \in [n] \setminus \{i\}$ and $1 \leq l' < l$. Now conditional on $\mathcal{G}$ and given that $j \in \text{rmargin}(G_m)$, for $m \leq l$, the event $j \in \bar{G}_{r'}$ has zero probability unless $r' \in \{m-1, m\}$.

Case 3.1. $r' = m-1$. In this case, $j \in \bar{G}_{m-1}$. Since, $j \in \text{rmargin}(G_m)$, the event $j \in \bar{G}_{m-1}$ depends only on the value of $\hat{f}_{j,m-1}$. Since, $m \leq l$, the random bit defining g_l is independent of the values of the random bits that determine $\hat{f}_{j,m-1}$. Therefore,

$$\mathsf{Pr}\left[g_l(i) = 1 \mid i \in \mathcal{S}_{l-1}, |\hat{f}_{i,l-1}| < Q_{l-1}, j \in \bar{G}_{r'}, \mathcal{G}\right] = \mathsf{Pr}\left[g_l(i) = 1 \mid \mathcal{G}\right] \ .$$

Arguing similarly, $\mathsf{Pr}\left[g_l(i) = 1 \mid i \in \mathcal{S}_l, j \in \bar{G}_{r'}, \mathcal{G}\right] = \mathsf{Pr}\left[g_l(i) = 1 \mid \mathcal{G}\right]$.

Therefore, it follows from Eqn. (99) that

$$\mathsf{Pr}\left[|\hat{f}_{i,l-1}| < Q_{l-1} \mid i \in \mathcal{S}_l, j \in \bar{G}_{r'}, \mathcal{G}\right] = \mathsf{Pr}\left[|\hat{f}_{i,l-1}| < Q_{l-1}, \mid i \in \mathcal{S}_{l-1}, j \in \bar{G}_{r'}\right] \tag{100}$$

Case 3.2. Suppose $r' = m$. Since $j \in \text{rmargin}(G_m)$, therefore, $j \in \bar{G}_m$ is equivalent to the event $\hat{f}_{j,m-1} < Q_{m-1}$ and $j \in \mathcal{S}_m$. If $m < l$, then the event $g_l(i) = 1$ is independent of the values of $\hat{f}_{j,m-1}$ and the event $j \in \mathcal{S}_m$. Hence the same conclusion as Eqn. (100) holds when $r' = m$ and $m < l$.

Now suppose $r' = m$ and $m = l$. Then, we have,

$$\begin{aligned}
&\mathsf{Pr}\left[g_l(i) = 1 \mid i \in \mathcal{S}_{l-1}, |\hat{f}_{i,l-1}| < Q_{l-1}, j \in \bar{G}_{r'}, \mathcal{G}\right] \\
&= \mathsf{Pr}\left[g_l(i) = 1 \mid j \in \bar{G}_{r'}, \mathcal{G}\right] \\
&= \mathsf{Pr}\left[g_l(i) = 1 \mid |\hat{f}_{j,l-1}| < Q_{l-1}, g_l(j) = 1, j \in \mathcal{S}_{l-1}, \mathcal{G}\right] \\
&= \mathsf{Pr}\left[g_l(i) = 1 \mid g_l(j) = 1, \mathcal{G}\right] \\
&= \mathsf{Pr}\left[g_l(i) = 1 \mid \mathcal{G}\right] \ .
\end{aligned}$$

Hence, Eqn. (100) continues to hold in this case as well. Thus in all cases, Eqn (100) holds.

Substituting this into Eqn. (98), we have,

$$\begin{aligned}
&\mathsf{Pr}\left[i \in \bar{G}_l, j \in \bar{G}_{r'} \mid \mathcal{G}\right] \\
&= q_{i,l-1,j,r'} \cdot \mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_{l-1}, \mathcal{G}\right] (2^{-l} \pm O(n^{-c}) \\
&\qquad + \mathsf{Pr}\left[|\hat{f}_{i,l-1}| < Q_{l-1} \mid i \in \mathcal{S}_{l-1}, j \in \bar{G}_{r'}, \mathcal{G}\right] \mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_l, \mathcal{G}\right] \mathsf{Pr}\left[i \in \mathcal{S}_l \mid \mathcal{G}\right] \\
&= \left(q_{i,l-1,j,r'} + 1 - (p_{i,l-1,j,r'} - q_{i,l-1,j,r'})\right) \mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_{l-1}, \mathcal{G}\right] (2^{-l} \pm O(n^{-c}) \\
&= (1 - p_{i,l-1,j,r'}) \mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_{l-1}, \mathcal{G}\right] (2^{-l} \pm O(n^{-c})
\end{aligned} \tag{101}$$

Multiplying Eqn. (97) by 2^{l-1} and Eqn. (101) by 2^l, we have for $r' \in \{m-1, m\}$ that

$$2^{l-1}\mathsf{Pr}\left[i \in \bar{G}_{l-1}, j \in \bar{G}_{r'} \mid \mathcal{G}\right] + 2^l \mathsf{Pr}\left[i \in \bar{G}_l, j \in \bar{G}_{r'} \mid \mathcal{G}\right] = \mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_{l-1}, \mathcal{G}\right] (1 \pm O(2^l n^{-c})) \tag{102}$$

The *LHS* of (102) can be equivalently written as $\sum_{r=0}^{L} 2^r \mathsf{Pr}\left[i \in \bar{G}_r, j \in \bar{G}_{r'} \mid \mathcal{G}\right]$, since, for $r \notin \{l-1, l\}$, $\mathsf{Pr}\left[i \in \bar{G}_r \mid \mathcal{G}\right] = 0$. Therefore,

$$\sum_{r=0}^{L} 2^r \mathsf{Pr}\left[i \in \bar{G}_r, j \in \bar{G}_{r'} \mid \mathcal{G}\right] = \mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_{l-1}, \mathcal{G}\right] (1 \pm O(2^l n^{-c})) \tag{103}$$

By Lemma 10, we have,

$$\sum_{r'=0}^{L} \mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_{l-1}, \mathcal{G}\right] = \sum_{r'=m-1}^{m} \mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_{l-1}, \mathcal{G}\right] = 1 \pm 2^m O(n^{-c})$$

Combining with Eqn. (103), we have,

$$\begin{aligned}\sum_{r'=0}^{L}\sum_{r=0}^{L} 2^r \mathsf{Pr}\left[i \in \bar{G}_r, j \in \bar{G}_{r'} \mid \mathcal{G}\right] &= \sum_{r'=0}^{L} \mathsf{Pr}\left[j \in \bar{G}_{r'} \mid i \in \mathcal{S}_{l-1}, \mathcal{G}\right](1 \pm O(2^l n^{-c})) \\ &= (1 \pm O(2^m n^{-c}))(1 \pm 2^l n^{-c})) = 1 \pm O((2^l + 2^m)n^{-c}) \ .\end{aligned}$$

□

F Application of Taylor polynomial estimator

Throughout the remainder of this section, let Y denote a code given by Corollary 5.

F.1 Preliminaries

Notation. We first partition the random seeds used by the algorithm by their functionality. For strings s and t, let $s \oplus t$ denote the string that is the concatenation of s and t.

Let $\bar{g}_l$ denote the random bit string representing the seed used to generate the hash function g_l, for $l \in \{0\} \cup [L]$, and let $\bar{g}$ denote the concatenation of the seed strings $\bar{g}_1 \oplus \bar{g}_2, \ldots \oplus \bar{g}_L$. For $l \in \{0\} \cup [L]$ and $j \in [s]$, let $\bar{h}_{HH,l,j}$ denote the random bit string used to generate the hash function corresponding to the jth hash table in the HH_l structure; let $\bar{h}_{HH,l}$ denote the concatenation of the random bitstrings $\oplus_{j\in[s]}\bar{h}_{HH,l}$ and $\bar{h}_{HH}$ denote the concatenation of the random bitstrings $\oplus_{l\in\{0,1,\ldots,L\}}\bar{h}_{HH,l}$. For $l \in \{0\} \cup [L]$ and $j \in [2s]$, let $\bar{h}_{lj}$ denote the random bit string used to generate the hash function h_{lj} in the $\textsc{tpest}_l$ structure. Let $\bar{h}_l$ denote the random bit string $\oplus_{j\in[2s]}\bar{h}_{lj}$ and let $\bar{h}$ denote the concatenation $\bar{h} = \oplus_{l\in\{0,1,\ldots,L\}}$. Let $\bar{\xi}_{HH,l,j}$ denote the random bit string used to generate the Rademacher family used by the jth table of the HH_l structure, for $l \in \{0, 1, \ldots, L\}$ and $j \in [s]$. Let $\bar{\xi}_{HH,l} = \oplus_{j\in[s]}\bar{\xi}_{HH,l,j}$ and let $\bar{\xi}_{HH} = \oplus_{l\in\{0,1,\ldots,L\}}\bar{\xi}_{HH,l}$. Let $\bar{\xi}_{lj}$ denote the random seed that generates the Rademacher variables $\{\xi_{lj}(k)\}_{k\in[n]}$ used by the jth table in $\textsc{tpest}_l$ structure, for $j \in [2s]$; let $\bar{\xi}_l = \oplus_{j\in[2s]}\bar{\xi}_{lj}$ and let $\bar{\xi} = \oplus_{l\in\{0,1,\ldots,L\}}\bar{\xi}_l$. Let $\bar{\zeta}$ denote the random bit string used to estimate F_2.

The full random seed string used to update and maintain the $\mathsf{Geometric\text{-}Hss}$ structure is $\bar{\zeta} \oplus \bar{g} \oplus \bar{h}_{HH} \oplus \bar{\xi}_{HH} \oplus \bar{h} \oplus \bar{\xi}$. In addition, during estimation, an n-dimensional random bit vector K is also used.

Note that the events in $\mathcal{G}$ are dependent only on $\bar{\zeta} \oplus \bar{g} \oplus \bar{h}_{HH} \oplus \bar{\xi}_{HH}$. This is further explained in the table below.

Event	Random bit string that determines the event
GOODF$_2$	$\bar{\zeta}$
NOCOLL	$\bar{h}$
GOODEST	$\bar{h}_{HH}$
SMALLRES	$\bar{g}$
ACCUEST	$\bar{g} \oplus \bar{h}_{HH}$
GOODFINALLEVEL	$\bar{g} \oplus \bar{h}_{HH}$
SMALLHH	$\bar{g} \oplus \bar{h}_{HH}$

F.2 Basic properties of the application of Taylor polynomial estimator: Proof of Lemma 13-Part I

For items $i, k \in [n]$ with $k \neq i$, hash table index $j \in [s]$ and $l \in [L] \cup \{0\}$, define the indicator variable u_{ikjl} to be 1 if $h_{lj}(i) = h_{lj}(k)$.

Proof of Lemma 13, parts (a), (b) and (e). Suppose $\mathcal{G}$ holds. The last statement of the lemma follows from GOODFINALLEVEL, which is a sub-event of $\mathcal{G}$.

Let $l = l_d(i) \in \{0\} \cup [L-1]$. By ACCUEST, $|\hat{f}_{il} - f_i| \leq \bar{\epsilon} T_l$. Since i is discovered at level l, $|\hat{f}_{il}| \geq Q_l = T_l - \bar{\epsilon} T_l$. So, $|f_i| \geq |\hat{f}_{il}| - \bar{\epsilon} T_l \geq Q_l - \bar{\epsilon} T_l = T_l - 2\bar{\epsilon} T_l$ and therefore,

$$\frac{|\hat{f}_i - f_i|}{|f_i|} \leq \frac{\bar{\epsilon} T_l}{(1 - 2\bar{\epsilon}) T_l} \leq \frac{1/(27p)}{(1 - 2/(27p))} < \frac{1}{26p}$$

since, $\bar{\epsilon} = (B/C)^{1/2} = 1/(27p)$ and $p \geq 2$. Therefore,

$$\frac{|\hat{f}_i - f_i|}{|\hat{f}_i|} \leq \frac{\bar{\epsilon} T_l}{(1 - \bar{\epsilon}) T_l} < \frac{1}{26p} \ .$$

This proves parts (a) and (e) of the lemma.

Let $j \in R_l(i)$ and $l_d(i) = l$. For $k \in [n]$, let y_{lk} be an indicator variable that is 1 if $k \in \mathcal{S}_l$ and is 0 otherwise. Then,

$$X_{ijl} = \sum_{k \in [n]} f_k \cdot y_{lk} \cdot \xi_{lj}(k) \cdot u_{ikjl} \cdot \xi_{lj}(i) \cdot \mathrm{sgn}(\hat{f}_i)$$

Since it is given that $l_d(i) = l$, it follows that

$$X_{ijl} = f_i \cdot \mathrm{sgn}(\hat{f}_i) + \sum_{k \in [n], k \neq i} f_k \cdot y_{lk} \cdot \xi_{lj}(k) \cdot u_{ikjl} \cdot \xi_{lj}(i) \cdot \mathrm{sgn}(\hat{f}_i) \ .$$

We now take expectations. Note that the events in $\mathcal{G}$ are independent of the Rademacher family random bits $\xi_{lj}(k)$. Also, the event $u_{ikjl} = 1$ depends only on $\bar{g} \oplus \bar{h}_l$ and the event $l_d(i) = l$ depends

only on $\bar{g} \oplus \bar{h}_{\mathsf{HH}}$. Therefore,

$$\begin{aligned}
&\mathbb{E}_{\bar{\xi}_{lj}}\left[X_{ijl} \mid l_d(i) = l, j \in R_l(i), \mathcal{G}\right] \\
&= f_i \cdot \operatorname{sgn}(\hat{f}_i) + \sum_{k \in [n]\setminus\{i\}} f_k \mathbb{E}_{\bar{\xi}_{lj}}\left[\xi_{lj}(k) \cdot \xi_{lj}(i) \cdot y_{lk} \cdot u_{ikjl} \mid l_d(i) = l, j \in R_l(i), \mathcal{G}\right] \\
&= f_i \cdot \operatorname{sgn}(\hat{f}_i) + \sum_{k \in \mathcal{S}_l} f_k \cdot \mathbb{E}_{\bar{\xi}_{lj}}\left[\xi_{lj}(k)\xi_{lj}(i) \mid y_{lk} = 1, u_{ikjl} = 1, j \in R_l(i), \mathcal{G}\right] \\
&\qquad\qquad \cdot \mathsf{Pr}\left[u_{ikjl} = 1, y_{lk} = 1 \mid j \in R_l(i), \mathcal{G}\right] \\
&= f_i \cdot \operatorname{sgn}(\hat{f}_i) + 0 \qquad (104)
\end{aligned}$$

since, $\xi_{lj}(k)$ and $\xi_{lj}(i)$ depend only on $\bar{\xi}_{lj}$ and is independent of the conditioning events. The expectation is zero by pair-wise independence and zero-expectation of the family $\{\xi_{lj}(s)\}_{s\in[n]}$.

Hence, Eqn. (104) becomes

$$\mathbb{E}_{\bar{\xi}_{lj}}\left[X_{ijl} \mid l_d(i) = l, j \in R_l(i), \mathcal{G}\right] = f_i \cdot \operatorname{sgn}(\hat{f}_i) = f_i \cdot \operatorname{sgn}(f_i) = |f_i| \qquad (105)$$

because, since, $l_d(i) = l$, $|\hat{f}_{il}| \geq (1-\bar{\epsilon})T_l$ and therefore, $\operatorname{sgn}(\hat{f}_i \pm \bar{\epsilon}T_l) = \operatorname{sgn}(\hat{f}_i)$, since, $\bar{\epsilon} = 1/(27p) < 1/2$. Since $\mathcal{G}$ holds, by ACCUEST we have,

$$\operatorname{sgn}(\hat{f}_i)\operatorname{sgn}(f_i) = \operatorname{sgn}(\hat{f}_i)\operatorname{sgn}(\hat{f}_i \pm \bar{\epsilon}T_l) = \operatorname{sgn}(\hat{f}_i)\operatorname{sgn}(\hat{f}_i) = 1$$

and therefore $\operatorname{sgn}(\hat{f}_i) = \operatorname{sgn}(f_i)$. Hence Eqn. (105) holds.

□

F.3 Expectation of $\bar{\vartheta}_i$

Proof of Lemma 14. By Lemma 13, we have,

$$\mathbb{E}_{\bar{\xi}_{lj}}\left[X_{ijl} \mid l_d(i) = l, j \in R_l(i), \mathcal{G}\right] = |f_i|$$

and therefore,

$$\mathbb{E}_{\bar{\xi}_l}\left[X_{ijl} \mid l_d(i) = l, j \in R_l(i), \mathcal{G}\right] = \mathbb{E}_{\bar{\xi}_l \setminus \bar{\xi}_{lj}}\left[\mathbb{E}_{\bar{\xi}_{lj}}\left[X_{ijl} \mid l_d(i) = l, j \in R_l(i), \mathcal{G}\right]\right] = |f_i| \ .$$

By Lemma 13 part (a), if i is discovered at level l, then, $|\hat{f}_{il} - f_i| \leq \frac{|f_i|}{26p}$.

Since NOCOLLISION holds as a sub-event of $\mathcal{G}$, $|R_l(i)| \geq s$. Let $\{j_1, j_2, \ldots, j_s\}$ be any s-subset of $R_l(i)$ such that $1 \leq j_1 < j_2 < \ldots < j_s \leq 2s$ and $y \in Y$ be a code with $\pi_y : [k] \to [k]$ being a random permutation. Let $y = (y_1, y_2, \ldots, y_k)$ be the k-dimensional increasing sequence $1 \leq y_1 < y_2 < \ldots < y_k \leq 2s$ representing the k non-zero positions in the s-dimensional bit vector y. Then, $\vartheta_{iyl} = \sum_{v=0}^{k} \binom{p}{v} |\hat{f}_i|^{p-v} \prod_{r=1}^{v} (X_{i,j_{y_{\pi(r)}},l} - |\hat{f}_i|)$. Therefore, each $j_{y_{\pi(r)}} \in R_l(i)$, for $1 \leq r \leq k$.

$$\begin{aligned}
&\mathbb{E}_{\bar{\xi}_l}\left[\vartheta_{iyl} \mid l_d(i) = l, \mathcal{G}\right] \\
&= \sum_{v=0}^{k} \binom{p}{v} |\hat{f}_i|^{p-v} \prod_{r=1}^{v} \left(\mathbb{E}_{\bar{\xi}_l}\left[X_{i,j_{y_{\pi(r)}},l} \mid l_d(i) = l, \mathcal{G}, j_{y_{\pi(r)}} \in R_l(i)\right] - |\hat{f}_i|\right) \\
&= \sum_{v=0}^{k} \binom{p}{v} |\hat{f}_i|^{p-v} \prod_{r=1}^{v} \left(|f_i| - |\hat{f}_i|\right)
\end{aligned}$$

which by Corollary 2 is bounded above as follows.

$$\begin{aligned}\left|\mathbb{E}_{\bar{\xi}_l}\left[\vartheta_{iyl} \mid l_d(i)=l,\mathcal{G}\right]-|f_i|^p\right| &\leq \left(\frac{\alpha}{1-\alpha}\right)^{k+1}\left(\frac{|f_i|}{k+1}\right)^p \\ &\leq \left(\frac{1/(26p)}{1-1/(26p)}\right)^{k+1}|f_i|^p \\ &\leq (25p)^{-k-1}|f_i|^p \leq n^{-4000p}|f_i|^p\end{aligned}$$

since $k \geq 1000\log(n)$.

Since, $\mathbb{E}\left[\bar{\vartheta}_i\right] = \mathbb{E}\left[\vartheta_{iyl}\right]$ for each $y \in Y$ and random permutation π_y, the lemma follows. Additionally, if p is integral then, $\mathbb{E}\left[\bar{\vartheta}_i\right] = \mathbb{E}\left[\vartheta_{il}\right] = |f_i|^p$.

□

F.3.1 Probability that two items collide conditional on the event NOCOLLISION

We first prove a lemma that bounds the probability that two distinct items collide under a hash function h_{lj} conditional on j being in $R_l(i)$.

Lemma 47. *Let $l_d(i) = l$, $k \in \mathcal{S}_l$ and $k \notin \widehat{\text{TOPK}}(C_l)$ and $i \neq k$. If the degree of independence of the hash family from which the hash functions h_{lj} are drawn is at least 11, then,*

$$\begin{aligned}&1.\ \mathsf{Pr}\left[u_{ikjl}=1 \mid l_d(i)=l, j\in R_l(i), k\in\mathcal{S}_l, k\notin\widehat{\text{TOPK}}(C_l)\right] \\ &\quad\in\left(1-\frac{1}{16C_l}\right)^{C_l-0.5\mp0.5} \pm 2\binom{C_l}{t-1}\left(\frac{1}{16C_l}\right)^{t-1},\ \textit{and,} \\ &2.\ \mathsf{Pr}\left[u_{ikjl}=1 \mid l_d(i)=l, j\in R_l(i), k\in\mathcal{S}_l, k\notin\widehat{\text{TOPK}}(C_l),\mathcal{G}\right] \\ &\quad\in\left(1-\frac{1}{16C_l}\right)^{C_l-1} \pm 2\binom{C_l}{t-1}\left(\frac{1}{16C_l}\right)^{t-1} \pm O(n^{-c})\ .\end{aligned}$$

Proof. Since $u_{ikjl} = 1$ is equivalent to $h_{lj}(i) = h_{lj}(k)$, we have,

$$\begin{aligned}&\mathsf{Pr}_t\left[u_{ikjl}=1 \mid l_d(i)=l, j\in R_l(i), k\in\mathcal{S}_l, k\notin\widehat{\text{TOPK}}(C_l)\right] \\ &=\left(\frac{\mathsf{Pr}_t\left[j\in R_l(i) \mid u_{ikjl}=1, l_d(i)=l, k\in\mathcal{S}_l, k\notin\widehat{\text{TOPK}}(C_l)\right]}{\mathsf{Pr}_t\left[j\in R_l(i) \mid l_d(i)=l, k\in\mathcal{S}_l, k\notin\widehat{\text{TOPK}}(C_l)\right]}\right) \\ &\quad\cdot\mathsf{Pr}_t\left[u_{ikjl}=1 \mid l_d(i)=l, k\in\mathcal{S}_l, k\notin\widehat{\text{TOPK}}(C_l)\right]\ . \qquad (106)\end{aligned}$$

First,

$$\begin{aligned}&\mathsf{Pr}_t\left[u_{ikjl}=1 \mid l_d(i)=l, k\in\mathcal{S}_l, k\notin\widehat{\text{TOPK}}(C_l)\right] \\ &\quad=\mathsf{Pr}_t\left[u_{ikjl}=1\right]=\frac{1}{16C_l}\pm\left(\frac{e}{16t}\right)^t \qquad (107)\end{aligned}$$

since, the event $u_{ikjl} = 1$ depends solely on $\bar{h}_{lj}$ and is independent of the events $k \in \mathcal{S}_l$ and $k \notin \widehat{\text{TOPK}}C_l$.

Secondly,

$$\begin{aligned}
&\mathsf{Pr}_t\left[j \in R_l(i) \mid u_{ikjl}=1, l_d(i)=l, k\in\mathcal{S}_l, k\notin \widehat{\textsc{Topk}}(C_l)\right]\\
&=\mathsf{Pr}_t\left[\forall i'\in \widehat{\textsc{Topk}}(C_l)\setminus\{i\}(h_{lj}(i')\neq h_{lj}(i)) \mid u_{ikjl}=1, l_d(i)=l, k\in\mathcal{S}_l, k\notin\widehat{\textsc{Topk}}(C_l)\right]\\
&=\frac{\mathsf{Pr}_t\left[\left(\forall i'\in \widehat{\textsc{Topk}}(C_l)\setminus\{i\}(h_{lj}(i')\neq h_{lj}(i))\right) \text{ and } u_{ikjl}=1 \mid l_d(i)=l, k\in\mathcal{S}_l, k\notin\widehat{\textsc{Topk}}(C_l)\right]}{\mathsf{Pr}_t\left[u_{ikjl}=1 \mid l_d(i)=l, k\in\mathcal{S}_l, k\notin\widehat{\textsc{Topk}}(C_l)\right]}\\
&=\frac{\mathsf{Pr}_t\left[\left(\forall i'\in \widehat{\textsc{Topk}}(C_l)\setminus\{i\}(h_{lj}(i')\neq h_{lj}(i))\right) \text{ and } u_{ikjl}=1 \mid l_d(i)=l, k\in\mathcal{S}_l, k\notin\widehat{\textsc{Topk}}(C_l)\right]}{\mathsf{Pr}_t\left[u_{ikjl}=1\right]}.
\end{aligned} \tag{108}$$

$u_{ikjl}=1$ is a function solely of $\bar{h}_{lj}$ and it is independent of the events $l_d(i)=l$ and $k\notin\widehat{\textsc{Topk}}(C_l)$. Hence, the denominator term in Eqn. (108) is simply $\mathsf{Pr}_t\left[u_{ikjl}=1\right]$.

Consider the numerator of Eqn. (108). Let $A=\widehat{\textsc{Topk}}(C_l)$, $|A|=k$. Then,

$$\begin{aligned}
&\mathsf{Pr}\left[\left(\forall i'\in \widehat{\textsc{Topk}}(C_l)\setminus\{i\}(h_{lj}(i')\neq h_{lj}(i))\right) \text{ and } u_{ikjl}=1 \mid l_d(i)=l, k\in\mathcal{S}_l, k\notin\widehat{\textsc{Topk}}(C_l)\right]\\
&=\sum_{A\subset[n],|A|=C_l}\mathsf{Pr}_t\left[\left(\forall i'\in A\setminus\{i\}(h_{lj}(i')\neq h_{lj}(i))\right), u_{ikjl}=1 \mid l_d(i)=l, k\in\mathcal{S}_l, k\notin A, \widehat{\textsc{Topk}}(C_l)=A\right]\\
&\qquad\qquad\cdot\mathsf{Pr}_{\bar{g}\oplus\bar{h}_{HH}}\left[\widehat{\textsc{Topk}}(C_l)=A \mid l_d(i)=l, k\in\mathcal{S}_l, k\notin A\right]\\
&=\sum_{\substack{A\subset[n]\\|A|=C_l}}\mathsf{Pr}_t\left[\left(\forall i'\in A\setminus\{i\}(h_{lj}(i')\neq h_{lj}(i))\right), u_{ikjl}=1\right]\mathsf{Pr}\left[\widehat{\textsc{Topk}}(C_l)=A \mid l_d(i)=l, k\in\mathcal{S}_l, k\notin A\right]
\end{aligned} \tag{109}$$

since for a fixed A, the event $\{\forall i'\in A\setminus\{i\}(h_{lj}(i')\neq h_{lj}(i)) \text{ and } u_{ikjl}=1\}$ is independent of the events $l_d(i)=l, k\in\mathcal{S}_l$ and $k\notin A$.

We now estimate the probability $\mathsf{Pr}_t\left[(\forall i'\in A\setminus\{i\}(h_{lj}(i')\neq h_{lj}(i))), u_{ikjl}=1\right]$. The event $\left\{\forall i'\in A\setminus\{i\}(h_{lj}(i')\neq h_{lj}(i)), u_{ikjl}=1\right\}$ is equivalent to $\left(\neg\bigvee_{i'\in A\setminus\{i\}}(u_{ii'jl}=1)\right)\wedge(u_{ikjl}=1)$. Therefore, by inclusion-exclusion, we have,

$$\begin{aligned}
&\mathsf{Pr}_t\left[\left(\neg\bigvee_{i'\in A\setminus\{i\}}(u_{ii'jl}=1)\right)\wedge(u_{ikjl}=1)\right]\\
&=\mathsf{Pr}_t\left[\neg\bigvee_{i'\in A\setminus\{i\}}(u_{ii'jl}=1) \mid u_{ikjl}=1\right]\mathsf{Pr}_t\left[u_{ikjl}=1\right]\\
&=\left(1-\mathsf{Pr}_t\left[\bigvee_{i'\in A\setminus\{i\}}(u_{ii'jl}=1) \mid u_{ikjl}=1\right]\right)\mathsf{Pr}_t\left[u_{ikjl}=1\right]
\end{aligned}$$

Following the inclusion-exclusion arguments as in Lemma 42 and using the notation that that $P[\cdot]$

denotes the probability measure assuming full-independence of the same hash family, we have,

$$\begin{aligned}
&\left|\left(1-\mathsf{Pr}_t\left[\bigvee_{i'\in A\setminus\{i\}}(u_{ii'jl}=1)\mid u_{ikjl}=1\right]\right)-\left(1-P\left[\bigvee_{i'\in A\setminus\{i\}}(u_{ii'jl}=1)\mid u_{ikjl}=1\right]\right)\right| \\
&\leq 2\sum_{\{i_1,i_2,\ldots,i_{t-1}\}\subset A\setminus\{i\}} P\left[\bigwedge_{r=1}^{t-1}u_{ii_rjl}=1\mid u_{ikjl}=1\right] \\
&\leq 2\binom{C_l}{t-1}\left(\frac{1}{16C_l}\right)^{t-1}
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathsf{Pr}_t\left[\left(\forall i'\in A\setminus\{i\}(h_{lj}(i')\neq h_{lj}(i))\right),u_{ikjl}=1\right] \\
&=\left(1-\mathsf{Pr}_t\left[\bigvee_{i'\in A\setminus\{i\}}(u_{ii'jl}=1)\mid u_{ikjl}=1\right]\right)\mathsf{Pr}\left[u_{ikjl}=1\right] \\
&=\left(\left(1-P\left[\bigvee_{i'\in A\setminus\{i\}}(u_{ii'jl}=1)\mid u_{ikjl}=1\right]\right)\pm 2\binom{C_l}{t-1}\left(\frac{1}{16C_l}\right)^{t-1}\right)\mathsf{Pr}\left[u_{ikjl}=1\right] \\
&=\left(\left(1-\frac{1}{16C_l}\right)^{C_l-\mathbf{1}_{i\notin A}}\pm 2\binom{C_l}{t-1}\left(\frac{1}{16C_l}\right)^{t-1}\right)\mathsf{Pr}\left[u_{ikjl}=1\right]
\end{aligned}$$

Now, $C_l-\mathbf{1}_{i\notin A}\in C_l-0.5\mp 0.5$.

Substituting in Eqn. (109), we have,

$$\begin{aligned}
&\mathsf{Pr}_t\left[\left(\forall i'\in\widehat{\textsc{Topk}}(C_l)\setminus\{i\}(h_{lj}(i')\neq h_{lj}(i))\right)\text{ and }u_{ikjl}=1\mid l_d(i)=l,k\in\mathcal{S}_l,k\notin\widehat{\textsc{Topk}}(C_l)\right] \\
&=\sum_{A\subset[n],|A|=k}\mathsf{Pr}_t\left[\left(\forall i'\in A\setminus\{i\}(h_{lj}(i')\neq h_{lj}(i))\right)\text{ and }u_{ikjl}=1\right] \\
&\qquad\cdot\mathsf{Pr}_{\bar{g}\oplus\bar{h}_{HH}}\left[\widehat{\textsc{Topk}}(C_l)=A\mid l_d(i)=l,k\in\mathcal{S}_l,k\notin A\right] \\
&\in\left(\left(1-\frac{1}{16C_l}\right)^{C_l-0.5\mp0.5}\pm 2\binom{C_l}{t-1}\left(\frac{1}{16C_l}\right)^{t-1}\right)\mathsf{Pr}\left[u_{ikjl}=1\right] \\
&\qquad\cdot\sum_{A\subset[n],|A|=k}\mathsf{Pr}_{\bar{g}\oplus\bar{h}_{HH}}\left[\widehat{\textsc{Topk}}(C_l)=A\mid l_d(i)=l,k\in\mathcal{S}_l,k\notin A\right] \\
&=\left(\left(1-\frac{1}{16C_l}\right)^{C_l-0.5\mp0.5}\pm 2\binom{C_l}{t-1}\left(\frac{1}{16C_l}\right)^{t-1}\right)\mathsf{Pr}\left[u_{ikjl}=1\right] \qquad (110)
\end{aligned}$$

Substituting in Eqn (108), we have,

$$\begin{aligned}
&\mathsf{Pr}_t\left[j \in R_l(i) \mid u_{ikjl}=1, l_d(i)=l, k \in \mathcal{S}_l, k \notin \widehat{\textsc{Topk}}(C_l)\right] \\
&= \left(\frac{1}{\mathsf{Pr}_t\left[u_{ikjl}=1\right]}\right) \mathsf{Pr}_t\left[\left(\forall i' \in \widehat{\textsc{Topk}}(C_l) \setminus \{i\}(h_{lj}(i') \neq h_{lj}(i))\right) \text{ and } u_{ikjl}=1\right. \\
&\qquad\qquad \left. \mid l_d(i)=l, k \in \mathcal{S}_l, k \notin \widehat{\textsc{Topk}}(C_l)\right] \\
&= \left(1-\frac{1}{16C_l}\right)^{C_l-0.5\mp0.5} \pm 2\binom{C_l}{t-1}\left(\frac{1}{16C_l}\right)^{t-1} \qquad (111)
\end{aligned}$$

In a similar manner, we can show that

$$\begin{aligned}
&\mathsf{Pr}_t\left[j \in R_l(i) \mid l_d(i)=l, k \in \mathcal{S}_l, k \notin \widehat{\textsc{Topk}}(C_l)\right] \\
&= \left(1-\frac{1}{16C_l}\right)^{C_l-0.5\mp0.5} \pm 2\binom{C_l}{t}\left(\frac{1}{16C_l}\right)^{t} \qquad (112)
\end{aligned}$$

Substituting Eqns. (111), (112) and (F.3.1) in Eqn. (106), we have,

$$\begin{aligned}
&\mathsf{Pr}_t\left[u_{ikjl}=1 \mid l_d(i)=l, j \in R_l(i), k \in \mathcal{S}_l, k \notin \widehat{\textsc{Topk}}(C_l)\right] \\
&= \left(\frac{\mathsf{Pr}_t\left[j \in R_l(i) \mid u_{ikjl}=1, l_d(i)=l, k \in \mathcal{S}_l, k \notin \widehat{\textsc{Topk}}(C_l)\right]}{\mathsf{Pr}_t\left[j \in R_l(i) \mid l_d(i)=l, k \in \mathcal{S}_l, k \notin \widehat{\textsc{Topk}}(C_l)\right]}\right) \\
&\qquad \cdot \mathsf{Pr}_t\left[u_{ikjl}=1 \mid l_d(i)=l, k \in \mathcal{S}_l, k \notin \widehat{\textsc{Topk}}(C_l)\right] \\
&= \left(\frac{\left(1-\frac{1}{16C_l}\right)^{C_l-0.5\mp0.5} \pm 2\binom{C_l}{t-1}\left(\frac{1}{16C_l}\right)^{t-1}}{\left(1-\frac{1}{16C_l}\right)^{C_l-0.5\pm0.5} \mp 2\binom{C_l}{t}\left(\frac{1}{16C_l}\right)^{t}}\right) \cdot \left(\frac{1}{16C_l} \pm \left(\frac{e}{16t}\right)^t\right) \qquad (113)
\end{aligned}$$

For $t=11$, the above ratio is bounded by $\left(\frac{1\pm10^{-16}}{16C_l}\right)$.

Conditioning with respect to $\mathcal{G}$, by Fact 43, the above probability may change by n^{-c}. Also, conditioned on $\mathcal{G}$, we have that $l_d(i)=l$ implies that $i \in \widehat{\textsc{Topk}}(C_l)$. Hence,

$$\begin{aligned}
&\mathsf{Pr}_t\left[j \in R_l(i) \mid u_{ikjl}=1, l_d(i)=l, k \in \mathcal{S}_l, k \notin \widehat{\textsc{Topk}}(C_l), \mathcal{G}\right] \\
&= \left(1-\frac{1}{16C_l}\right)^{C_l-1} \pm 2\binom{C_l}{t-1}\left(\frac{1}{16C_l}\right)^{t-1} \pm n^{-c} \ .
\end{aligned}$$

Proceeding similarly as in Eqn. (113), we have,

$$\begin{aligned}
&\mathsf{Pr}_t\left[u_{ikjl}=1 \mid l_d(i)=l, j\in R_l(i), k\in\mathcal{S}_l, k\notin \widehat{\textsc{Topk}}(C_l),\mathcal{G}\right]\\
&=\left(\frac{\mathsf{Pr}_t\left[j\in R_l(i)\mid u_{ikjl}=1, l_d(i)=l, k\in\mathcal{S}_l, k\notin \widehat{\textsc{Topk}}(C_l),\mathcal{G}\right]}{\mathsf{Pr}_t\left[j\in R_l(i)\mid l_d(i)=l, k\in\mathcal{S}_l, k\notin \widehat{\textsc{Topk}}(C_l),\mathcal{G}\right]}\right)\\
&\qquad\qquad\cdot\mathsf{Pr}_t\left[u_{ikjl}=1\mid l_d(i)=l, k\in\mathcal{S}_l, k\notin \widehat{\textsc{Topk}}(C_l),\mathcal{G}\right]\\
&=\left(\frac{\left(1-\frac{1}{16C_l}\right)^{C_l-0.5\mp0.5}\pm 2\binom{C_l}{t-1}\left(\frac{1}{16C_l}\right)^{t-1}\pm n^{-c}}{\left(1-\frac{1}{16C_l}\right)^{C_l-0.5\pm0.5}\mp 2\binom{C_l}{t}\left(\frac{1}{16C_l}\right)^{t}\mp n^{-c}}\right)\cdot\left(\frac{1}{16C_l}\pm\left(\frac{e}{16t}\right)^t\pm n^{-c}\right)
\end{aligned}$$

For $t=11$, the above ratio is bounded by $\left(\frac{1\pm10^{-16}}{16C_l}\right)$.

□

F.4 Basic properties of the application of Taylor polynomial estimator: Proof of Lemma 13-Part II

We now complete the proofs of the remaining parts of Lemma 13.

Proof of Lemma 13, parts (c), (d) and (f). Recall that y_{lk} is an indicator variable that is 1 iff $k\in\mathcal{S}_l$. Given that $i\in\mathcal{S}_l$ the random variable X_{ijl} is defined as

$$X_{ijl}=(f_i+\sum_{k\neq i}f_k\cdot u_{ikjl}\cdot\xi_{lj}(k)\cdot\xi_{lj}(i)\cdot y_{lk})\mathrm{sgn}(\hat{f}_i)\ .$$

As shown in the proof of Lemma 14, $\mathbb{E}_{\bar{\xi}_{lj}}\left[X_{ijl}\mid j\in R_l(i), l_d(i)=l,\mathcal{G}\right]=|f_i|$. Further,

$$\begin{aligned}
&\mathbb{E}\left[X_{ijl}^2\mid j\in R_l(i), l_d(i)=l,\mathcal{G}\right]\\
&=\mathbb{E}_{\bar{h}_{HH,l}\oplus\bar{h}_{lj}\oplus\bar{g}}\left[\mathbb{E}_{\bar{\xi}_{lj}}\left[X_{ijl}^2\mid j\in R_l(i), l_d(i)=l,\mathcal{G}\right]\right]\\
&=f_i^2+\mathbb{E}_{\bar{h}_{HH,l}\oplus\bar{h}_{lj}\oplus\bar{g}}\left[\sum_{k\in[n]\setminus\{i\}}f_k^2\cdot u_{ijkl}\cdot y_{lk}\mid j\in R_l(i), l_d(i)=l,\mathcal{G}\right]
\end{aligned}$$

since the expectation with respect to the Rademacher family of TPEST structure is independent of the random bits used to define $\mathcal{G}$ and $R_l(i)$.

Therefore,

$$\begin{aligned}
\sigma^2_{ijl} &= \mathsf{Var}_{\bar{\xi}_{lj}\oplus\bar{h}_{HH,l}\oplus\bar{h}_{lj}\oplus\bar{g}}\left[X_{ijl} \mid j \in R_l(i), l_d(i) = l, \mathcal{G}\right] \\
&= \mathbb{E}_{\bar{\xi}_{lj}\oplus\bar{h}_{HH,l}\oplus\bar{h}_{lj}\oplus\bar{g}}\left[X^2_{ijl} \mid j \in R_l(i), l_d(i) = l, \mathcal{G}\right] \\
&\qquad - \left(\mathbb{E}_{\bar{\xi}_{lj}\oplus\bar{h}_{HH,l}\oplus\bar{h}_{lj}\oplus\bar{g}}\left[X_{ijl} \mid j \in R_l(i), l_d(i) = l, \mathcal{G}\right]\right)^2 \\
&= f_i^2 + \mathbb{E}_{\bar{g}\oplus\bar{h}_{HH,l}\oplus\bar{h}_{lj}}\left[\sum_{k\in[n]\setminus\{i\}} f_k^2 \cdot u_{ikjl} \cdot y_{lk} \mid j \in R_l(i), l_d(i) = l, \mathcal{G}\right] - |f_i|^2 \\
&= \sum_{k\in[n]\setminus\{i\}} f_k^2 \cdot \mathsf{Pr}_{\bar{g}\oplus\bar{h}_{HH,l}\oplus\bar{h}_{lj}}\left[u_{ikjl} = 1 \mid j \in R_l(i), l_d(i) = l, k \in \mathcal{S}_l, \mathcal{G}\right] \\
&\qquad\qquad \cdot \mathsf{Pr}_{\bar{g}\oplus\bar{h}_{HH,l}\oplus\bar{h}_{lj}}\left[y_{lk} = 1 \mid j \in R_l(i), l_d(i) = l, \mathcal{G}\right] \quad . \qquad (114)
\end{aligned}$$

Now,

$$\begin{aligned}
&\mathsf{Pr}\left[u_{ikjl} = 1 \mid j \in R_l(i), l_d(i) = l, k \in \mathcal{S}_l, \mathcal{G}\right] \qquad (115) \\
&= \mathsf{Pr}\left[u_{ikjl} = 1, k \notin \widehat{\textsc{Topk}}(C_l) \mid j \in R_l(i), k \in \mathcal{S}_l, l_d(i) = l, \mathcal{G}\right] \\
&\qquad\qquad + \mathsf{Pr}\left[u_{ikjl} = 1, k \in \widehat{\textsc{Topk}}(C_l) \mid j \in R_l(i), l_d(i) = k, \mathcal{G}\right] \\
&= \mathsf{Pr}\left[u_{ikjl} = 1 \mid j \in R_l(i), k \in \mathcal{S}_l, k \notin \widehat{\textsc{Topk}}(C_l), l_d(i) = l, \mathcal{G}\right] \\
&\qquad\qquad \cdot \mathsf{Pr}\left[k \notin \widehat{\textsc{Topk}}(C_l) \mid j \in R_l(i), k \in \mathcal{S}_l, l_d(i) = l, \mathcal{G}\right] + 0 \\
&\leq \left(\frac{1 + 10^{-16}}{(16C_l)}\right) \cdot \mathsf{Pr}\left[k \notin \widehat{\textsc{Topk}}(C_l) \mid j \in R_l(i), k \in \mathcal{S}_l, l_d(i) = l, \mathcal{G}\right] \qquad (116)
\end{aligned}$$

by (Lemma 47, with $t = 11$.

Substituting in (114), we have that

$$\begin{aligned}
\sigma^2_{ijl} &\leq \sum_{k\in[n]\setminus\{i\}} f_k^2 \cdot \left(\frac{1 + 10^{-16}}{(16C_l)}\right) \\
&\qquad \cdot \mathsf{Pr}\left[k \notin \widehat{\textsc{Topk}}(C_l), k \in \mathcal{S}_l \mid j \in R_l(i), l_d(i) = l, \mathcal{G}\right] \\
&\leq \left(\frac{1 + 10^{-16}}{(16C_l)}\right) \sum_{k\in[n]\setminus\{i\}, k\in\mathcal{S}_l, k\notin\widehat{\textsc{Topk}}(C_l)} f_k^2 \cdot 1 \\
&= \left(\frac{1 + 10^{-16}}{(16C_l)}\right) F_2^{\text{res}}\left(\widehat{\textsc{Topk}}(C_l), l\right) \qquad (117)
\end{aligned}$$

It can be shown that, conditional on GOODEST,

$$F_2^{\text{res}}\left(\widehat{\textsc{Topk}}(C_l), l\right) \leq 9F_2^{\text{res}}\left(C_l, l\right)$$

(this is explicitly proved in [18]; variants appear in earlier works for e.g., [16, 13, 22]). Since, SMALLRES holds as a sub-event of $\mathcal{G}$, $F_2^{\text{res}}\left(C_l, l\right) \leq 1.5F_2^{\text{res}}\left((2\alpha)^l C\right)/2^{l-1}$. Therefore, Eqn. (117)

may be written as follows.

$$\begin{aligned}
\sigma_{ijl}^2 &\le \left(\frac{1+10^{-16}}{(16C_l)}\right) F_2^{\text{res}}\left(\widehat{\text{TOPK}}(C_l), l\right) \\
&\le \frac{9(1+10^{-16})F_2^{\text{res}}(C_l,l)}{16C_l} \quad \text{since, } (F_2^{\text{res}}\left(\widehat{\text{TOPK}}_l(C_l), l\right) \le 9F_2^{\text{res}}(C_l,l)) \text{ [18, 22]} \\
&\le \frac{9(1+10^{-16}))(1.5)F_2^{\text{res}}\left((2\alpha)^l C\right)}{C_l(16)2^{l-1}} \qquad (\mathcal{G} \text{ implies SMALLRES.}) \\
&\le \frac{9(1+10^{-16})(1.5)\hat{F}_2}{8(2\alpha)^l C} \\
&\le (17/10)(\bar{\epsilon}T_l)^2 \ .
\end{aligned}$$

This proves part (d) of Lemma 13.

Hence,

$$\eta_{ijl}^2 = |\hat{f}_{il} - f_i|^2 + \sigma_{ijl}^2 \le (\bar{\epsilon}T_l)^2 + (17/10)(\bar{\epsilon}T_l)^2 \le 2.7(\bar{\epsilon}T_l)^2 \ .$$

Since, i is discovered at level l, $|\hat{f}_{il}| \ge Q_l = T_l(1-\bar{\epsilon})$ and therefore, $|f_i| \ge Q_l - \bar{\epsilon}T_l = T_l(1-2\bar{\epsilon})$.

Hence, $\frac{|f_i|}{\eta_{ijl}} \ge \frac{T_l(1-2\bar{\epsilon})}{(\sqrt{2.7})\bar{\epsilon}T_l} \ge 15p$. Further, $\frac{|\hat{f}_i|}{\eta_{ijl}} \ge \frac{T_l(1-\bar{\epsilon})}{\sqrt{2.7}\bar{\epsilon}T_l} \ge 16p$. This proves parts (c) and (f). □

F.5 Taylor polynomial estimators are uncorrelated with respect to $\bar{\xi}$

Proof of Lemma 15. The expectations in this proof are only with respect to $\bar{\xi}$.

Consider $\mathbb{E}_{\bar{\xi}}\left[\bar{\vartheta}_{i'}\bar{\vartheta}_i\right]$. $\bar{\vartheta}_i$ and $\bar{\vartheta}_{i'}$ each use the TPEst structure at levels $l_d(i)$ and $l_d(i')$ respectively. If $l_d(i) \ne l_d(i')$, then the estimations are made from different structures and use independent random bits and therefore,

$$\mathbb{E}_{\bar{\xi}}\left[\bar{\vartheta}_i\bar{\vartheta}_{i'} \mid \hat{f}_i, \hat{f}_{i'}\mathcal{G}\right] = \mathbb{E}_{\bar{\xi}}\left[\bar{\vartheta}_i \mid \hat{f}_i, \mathcal{G}\right]\mathbb{E}_{\bar{\xi}}\left[\bar{\vartheta}_{i'} \mid \hat{f}_i, \mathcal{G}\right] \ .$$

Now suppose that $l_d(i) = l_d(i') = l$ (say). Then, $|\hat{f}_{il}| \ge Q_l$ and $|\hat{f}_{i'l}| \ge Q_l$. Since SMALLHH holds as a sub-event of $\mathcal{G}$, $\{i,i'\} \subset \{k : |\hat{f}_{kl}| \ge Q_l\} \subset \widehat{\text{TOPK}}(l, C_l)$. Therefore, by NOCOLL$_l$, the estimates $\{X_{ijl}\}_{j\in R_l(i)}$ and $\{X_{i'jl}\}_{j\in R_l(i')}$ are such that if $j \in R_l(i) \cap R_l(i')$, then, $h_{lj}(i) \ne h_{lj}(i')$. Let $q_1, q_2, \ldots, q_s$ be some permutation of the table indices in $R_l(i)$. Likewise let $q'_1, q'_2, \ldots, q'_s$ be a permutation of the table indices in $R_l(j)$. Then,

$$\begin{aligned}
&\mathbb{E}_{\bar{\xi}}\left[\vartheta_i\vartheta_{i'} \mid \hat{f}_i, \hat{f}_{i'}, \mathcal{G}\right] \\
&= \mathbb{E}_{\bar{\xi}}\left[\left(\sum_{v=0}^{k}\gamma_v(|\hat{f}_i|)\prod_{w=1}^{v}(X_{i,q_w,l} - |\hat{f}_i|)\right)\left(\sum_{v'=0}^{k}\gamma_{v'}(|\hat{f}_{i'}|)\prod_{w'=1}^{v'}(X_{i',q'_w,l} - |\hat{f}_{i'}|)\right)\right. \\
&\qquad\qquad \left.\Big| \hat{f}_i, \hat{f}_{i'}, \mathcal{G}\right] \\
&= \sum_{v,v'=0}^{k}\gamma_v(|\hat{f}_i|)\gamma_{v'}(|\hat{f}_{i'}|)\mathbb{E}_{\bar{\xi}}\left[\prod_{w=1}^{v}(X_{i,q_w,l} - |\hat{f}_i|)\prod_{w'=1}^{v'}(X_{i',q'_{w'},l} - |\hat{f}_{i'}|) \mid \hat{f}_i, \hat{f}_{i'}, \mathcal{G}\right] \qquad (118)
\end{aligned}$$

Consider $\mathbb{E}_{\bar{\xi}}\left[\prod_{w=1}^{v}(X_{i,q_w,l} - |\hat{f}_i|)\prod_{w'=1}^{v'}(X_{i',q'_{w'},l} - |\hat{f}_{i'}|) \mid \hat{f}_i, \hat{f}_{i'}, \mathcal{G}\right]$. For some $1 \le w' \le v'$, if $q'_{w'} \notin \{q_1, q_2, \ldots, q_w\}$, then, the random variable $X_{i',q'_{w'},l} - |\hat{f}_{i'}|$ uses only the random bits of $\xi_{lq'_{w'}}$

and is independent of the random bits $\{\xi_{l,q_w} \mid 1 \le w \le v\}$ used by any of the $X_{i,q_w,l}$, for $1 \le w \le v$. An analogous situation holds for any $1 \le w \le v$ such that $q_w \notin \{q'_1, \ldots, q'_{v'}\}$. Clearly, for distinct tables, j, j', $\mathbb{E}_{\bar{\xi}}\left[X_{i,j,l}X_{i',j',l}\right]$ is the product of the individual expectations, by independence of the seeds of the Rademacher families $\{\xi_{lj}(k)\}$ and $\{\xi_{lj'}(k)\}$. Therefore,

$$\begin{aligned}
&\mathbb{E}_{\bar{\xi}}\left[\prod_{w=1}^{v}(X_{i,q_w,l} - |\hat{f}_i|)\prod_{w'=1}^{v'}(X_{i',q'_{w'},l} - |\hat{f}_{i'}|) \mid \hat{f}_i, \hat{f}_{i'}, \mathcal{G}\right] \\
&= \prod_{w:q_w \notin \{q'_1,\ldots,q'_{v'}\}} \mathbb{E}_{\xi_{l,q_w}}\left[(X_{i,q_w,l} - |\hat{f}_i|) \mid \hat{f}_i, \mathcal{G}\right] \\
&\quad \cdot \prod_{w':q'_{w'} \notin \{q_1,\ldots,q_v\}} \mathbb{E}_{\xi_{l,q'_{w'}}}\left[(X_{i',q'_{w'},l} - |\hat{f}_{i'}|) \mid \hat{f}_{i'}, \mathcal{G}\right] \\
&\quad \cdot \prod_{j \in \{q_1,\ldots,q_v\} \cap \{q'_1,\ldots,q'_{w'}\}} \mathbb{E}_{\xi_{lj}}\left[(X_{ijl} - |\hat{f}_i|)(X_{i'jl} - |\hat{f}_{i'}|) \mid \hat{f}_i, \hat{f}_{i'}, \mathcal{G}\right]
\end{aligned}$$

We analyze $\mathbb{E}_{\xi_{lj}}\left[X_{ijl}X_{i'jl} \mid \hat{f}_{il}, \hat{f}_{i'l}\mathcal{G}\right]$.

$$\begin{aligned}
&\mathbb{E}_{\xi_{lj}}\left[X_{ijl}X_{i'jl} \mid \hat{f}_{il}, \hat{f}_{i'l}, \mathcal{G}\right] = \mathrm{sgn}(f_i)\mathrm{sgn}(f_j) \\
&\cdot \mathbb{E}_{\xi_{lj}}\left[\left(f_i + \xi_{lj}(i)\sum_{k\neq i} f_k \cdot \xi_{lj}(k) \cdot u_{ikjl}\right) \cdot \left(f_{i'} + \xi_{lj}(i')\sum_{k'\neq i'} f_{k'} \cdot \xi_{lj}(k') \cdot u_{i'k'jl}\right)\right. \\
&\qquad \left. \Big| \hat{f}_{il}, \hat{f}_{i'l}, \mathcal{G}\right] \qquad (119)
\end{aligned}$$

Suppose we use linearity of expectation to expand the product and take the expectation of the individual terms. The expectation of the terms of the form $\mathbb{E}_{\xi_{lj}}\left[\xi_{lj}(i)\xi_{lj}(k)u_{ikjl}\right] = 0$ since $i \neq k$ and the random variable u_{ikjl} is independent of ξ_{lj}. Similarly, $\mathbb{E}_{\xi_{lj}}\left[\xi_{lj}(i')\xi_{lj}(k')u_{i'k'jl}\right] = 0$. We also obtain a set of terms of the form $\mathbb{E}_{\xi_{lj}}\left[\xi_{lj}(i) \cdot \xi_{lj}(i') \cdot \xi_{lj}(k) \cdot \xi_{lj}(k') \cdot u_{ikjl} \cdot u_{i'k'jl}\right]$. Since, $j \in R_l(i) \cap R_l(i')$, $h_{lj}(i) \neq h_{lj}(i')$. Now $u_{ikjl} \cdot u_{i'k'jl} = 1$ only if $h_{lj}(i) = h_{lj}(k)$ and $h_{lj}(i') = h_{lj}(k')$. We conclude that $\{i, i', k, k'\}$ are all distinct, and by 4-wise independence of the $\{\xi_{lj}(u)\}_{1\le u\le n}$ family, $\mathbb{E}_{\xi_{lj}}\left[\xi_{lj}(i) \cdot \xi_{lj}(i') \cdot \xi_{lj}(k) \cdot \xi_{lj}(k') \cdot u_{ikjl} \cdot u_{i'k'jl}\right] = 0$. Therefore, Eqn. (119) becomes

$$\mathbb{E}_{\xi_{lj}}\left[X_{ijl}X_{i'jl} \mid \hat{f}_{il}, \hat{f}_{i'l}, \mathcal{G}\right] = |f_i||f_{i'}| = \mathbb{E}_{\xi_{lj}}\left[X_{ijl} \mid \hat{f}_{il}, \mathcal{G}\right]\mathbb{E}_{\xi_{lj}}\left[X_{ijl} \mid \hat{f}_{i'l}, \mathcal{G}\right] \ .$$

It follows that

$$\begin{aligned}
&\mathbb{E}_{\xi_{lj}}\left[(X_{ijl} - |\hat{f}_{il}|)(X_{i'jl} - |\hat{f}_{i'l}|) \mid \hat{f}_{il}, \hat{f}_{i'l}, \mathcal{G}\right] = (|f_i| - |\hat{f}_{il}|))(|f_{i'}| - |\hat{f}_{i'l}|) \\
&= \mathbb{E}_{\xi_{lj}}\left[X_{ijl} - |\hat{f}_{il}| \mid \hat{f}_{il}, \mathcal{G}\right]\mathbb{E}_{\xi_{lj}}\left[X_{i'jl} - |\hat{f}_{i'l}| \mid \hat{f}_{il}, \mathcal{G}\right] \ .
\end{aligned}$$

For $l_d(i) = l_d(i') = l$, (118) simplifies to

$$\mathbb{E}_{\xi_l}\left[\vartheta_i\vartheta_{i'} \mid \hat{f}_{il}, \hat{f}_{i'l}, \mathcal{G}\right] = \mathbb{E}_{\xi_l}\left[\vartheta_i \mid \hat{f}_{il}, \mathcal{G}\right]\mathbb{E}_{\xi_l}\left[\vartheta_{i'} \mid \hat{f}_{i'l}, \mathcal{G}\right] \ .$$

Thus, ϑ_i and $\vartheta_{i'}$ are uncorrelated in all cases.

Since, $\bar{\vartheta}_i$ is the average of the Taylor polynomial estimators ϑ_i for randomly chosen permutations, the variables $\bar{\vartheta}_i$ and $\bar{\vartheta}_{i'}$ are also uncorrelated in all cases, whether $l \neq l'$ or $l = l'$, that is,

$$\mathbb{E}_{\bar{\xi}}\left[\bar{\vartheta}_i\bar{\vartheta}_{i'} \mid \hat{f}_{il}, \hat{f}_{i'l'}, \mathcal{G}\right] = \mathbb{E}_{\xi_l}\left[\bar{\vartheta}_i \mid \hat{f}_{il}, \mathcal{G}\right]\mathbb{E}_{\xi_{l'}}\left[\bar{\vartheta}_{i'} \mid \hat{f}_{i'l'}, \mathcal{G}\right] \tag{120}$$

□

G Expectation and Variance of pth moment estimator

In this section, we analyze the expectation and variance of the estimator $\hat{F}_p$.

G.1 Expectation of the $\hat{F}_p$ estimator

Proof of Lemma 16. Define $\mathsf{level} : [n] \to \{0, 1, 2 \ldots, L+1\}$ to be the function that maps each item $i \in [n]$ to the index of the group it belongs to, that is,

$$\mathsf{level}(i) = \begin{cases} l & \text{if } i \in G_l \\ L+1 & \text{if } f_i = 0. \end{cases}$$

Then, by definition of the Y_i's,

$$\begin{aligned}
\mathbb{E}\big[\hat{F}_p \mid \mathcal{G}\big] &= \mathbb{E}\Big[\sum_{i\in[n]} Y_i \mid \mathcal{G}\Big] \\
&= \sum_{l=0}^{L}\sum_{i\in G_l}\sum_{l'=0}^{L} 2^{l'}\mathbb{E}\left[z_{il'}\bar{\vartheta}_i \mid \mathcal{G}\right] \\
&= \sum_{l=0}^{L}\sum_{i\in G_l}\sum_{l'=0}^{L} 2^{l'}\mathbb{E}\left[\bar{\vartheta}_i \mid i \in \bar{G}_{l'}, \mathcal{G}\right]\mathsf{Pr}\left[i \in \bar{G}_{l'} \mid \mathcal{G}\right] \\
&= \sum_{l=0}^{L}\sum_{i\in G_l}\sum_{l'=0}^{L} 2^{l'}|f_i|^p(1 \pm n^{-4000p})\mathsf{Pr}\left[i \in \bar{G}_{l'} \mid \mathcal{G}\right], \quad \text{by Lemma 14} \\
&= \sum_{l=0}^{L}\sum_{i\in G_l}|f_i|^p(1 \pm n^{-4000p})\sum_{l'=0}^{L} 2^{l'}\mathsf{Pr}\left[i \in \bar{G}_{l'} \mid \mathcal{G}\right] \\
&= \sum_{l=0}^{L}\sum_{i\in G_l}|f_i|^p(1 \pm n^{-4000p})(1 \pm O(2^{\mathsf{level}(i)}n^{-c})), \quad \text{by Lemma 8} \\
&= F_p(1 \pm 2^{L+1}n^{-c}) \ .
\end{aligned}$$

Let $C = K'n^{1-2/p}$ where $K' = \dfrac{(27p)^2\epsilon^{-2}}{\min(\epsilon^{4/p-2}, \log n)}$, as given in Figure 2.

Since $\alpha = 1 - (1-2/p)(0.01) > 0.99$,

$$L = \lceil \log_{2\alpha}(n/C) \rceil \leq 1 + \log_{1.98}(n/C) \leq 1 + (1.02)\log_2(n/C) \leq 1 + (1.02)\log_2(n^{2/p}/K') \ .$$

Hence,

$$2^L \leq 2\left(\frac{n^{(2/p)}}{K'}\right)^{1.02}$$

and so $O(2^{L+1}n^{-c}) = O(n^{-(c-2)})$ proving the lemma. □

G.2 Variance of Y_i

In this section, we calculate $\mathsf{Var}[Y_i]$. For sake of completeness we first present proofs of some identities stated in Eqn. (8).

Fact 48. *For any $p \geq q$, $F_q \leq n^{1-q/p}F_p^{q/p}$. In particular, $F_2 \leq n^{1-2/p}F_p^{2/p}$ for any $p \geq 2$.*

Proof. Let X be a random variable that takes the value $|f_i|^q$ with probability $1/n$, for $i \in [n]$. Then,

$$\mathbb{E}[X] = \frac{F_q}{n} \ .$$

By Jensen's inequality, for any function f that is convex over the support of X, $\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$. Choose $f(t) = t^{p/q}$. Since $p \geq q$ and the support of X is $\mathbb{R}^{\geq 0}$, $f(t)$ is convex in this range. Therefore, $\mathbb{E}[f(X)] = \frac{F_p}{n}$. By Jensen's inequality applied to f, we have,

$$\left(\frac{F_q}{n}\right)^{p/q} \leq \frac{F_p}{n}, \quad \text{or,} \quad F_q \leq n^{1-q/p}F_p^{q/p} \ .$$

□

In the following proofs, we will use the notion that the *sample group of an item is consistent with the frequency of the item* to mean that if $i \in G_l$ and i is sampled into $\bar{G}_r$, then, l and r are related as given by Lemma 8, conditional on $\mathcal{G}$. (For e.g., if $i \in \text{lmargin}(G_l)$, then, $r \in \{l, l+1\}$, if $i \in \text{mid}(G_l)$, then, $r = l$, and if $i \in \text{rmargin}(G_l)$, then, $r \in \{l-1, l\}$).

Proof of Lemma 17. For this proof, assume that $\mathcal{G}$ holds.

Case 1: $i \in \text{mid}(G_0)$. Then $i \in \bar{G}_0$ with probability 1 and $l_d(i) = 0$. Therefore,

$$Y_i = \sum_{l=0}^{L-1} 2^l \cdot z_{il} \cdot \bar{\vartheta}_{il} = \bar{\vartheta}_{i0}$$

since, $z_{i0} = 1$ and $z_{il} = 0$ for $l > 0$. Let $\bar{\vartheta}_i$ denote $\bar{\vartheta}_{i0}$. Therefore, $\mathsf{Var}[Y_i \mid \mathcal{G}] = \mathsf{Var}\left[\bar{\vartheta}_i \mid \mathcal{G}\right]$.

From Figure 2, we have, $C = (27p)^2 B \geq (27p)^2 K\epsilon^{-2}n^{1-2/p}/\log(n)$. Since the estimator $\bar{\vartheta}_i$ uses the TPEST structure at level 0, by Lemma 13 (part (b)), we have, $\mu = \mathbb{E}[X_{ij0} \mid \mathcal{G}] = |f_i|$ and by

part (iv) of the same lemma, $\eta_{ij0}^2 \leq (2.7)\hat{F}_2/C$, for each $j \in R_0(i)$. Therefore, by Lemma 6,

$$\begin{aligned}
\mathsf{Var}\left[\bar{\vartheta}_i \mid i \in \bar{G}_0, \mathcal{G}\right] &\leq \left(\frac{(0.288)p^2}{k}\right)|f_i|^{2p-2}\eta_{ij0}^2 \\
&\leq \left(\frac{(0.288)p^2|f_i|^{2p-2}}{(1000)(\log n)}\right)\left(\frac{2.7\hat{F}_2}{C}\right) \\
&\leq \left(\frac{(0.288)p^2|f_i|^{2p-2}}{(1000)(\log n)}\right)\left(\frac{(2.7)(1.0005)F_2}{(27)p^2K\epsilon^{-2}n^{1-2/p}/(\log(n))}\right) \\
&\leq \frac{(0.3)\epsilon^2|f_i|^{2p-2}F_p^{2/p}}{(10)^4K} \ . \qquad (121)
\end{aligned}$$

where, the last step uses the fact that $F_2 \leq F_p^{2/p}n^{1-2/p}$, for $p > 2$ from (8), and that $\hat{F}_2 \leq (1 + 0.001/(2p))F_2$.

Case 2: $i \in lmargin(G_0) \cup_{r=1}^{L} G_r$. If $i \in G_l$, then, $l_d(i) \in \{l, l-1\}$ and if $i \in \bar{G}_r$ then $l - 1 \leq r \leq l + 1$. By Lemma 13, $\eta_{ijl_d(i)} \leq |f_i|/(15p)$ for $j \in R_l(i)$. From Lemma 6, we have,

$$\mathsf{Var}\left[\bar{\vartheta}_i \mid i \in \bar{G}_r, \mathcal{G}\right] = \left(\frac{(0.288)p^2}{k}\right)|f_i|^{2p-2}\eta_{ijl_d(i)}^2 \leq \frac{|f_i|^{2p}}{(750)k} \ .$$

Hence,

$$\begin{aligned}
\mathsf{Var}\left[Y_i \mid \mathcal{G}\right] &= \mathsf{Var}\left[\sum_{r=0}^{L} 2^r\bar{\vartheta}_i z_{ir} \mid \mathcal{G}\right] \\
&= \sum_{r=0}^{L} 2^{2r}\mathsf{Var}\left[\bar{\vartheta}_i z_{ir}\right] + \sum_{\substack{0\leq r,r'\leq L \\ r\neq r'}} 2^{r+r'}\mathsf{Cov}\left(\bar{\vartheta}_i z_{ir}, \bar{\vartheta}_i z_{ir'}\right) \\
&= \sum_{r=0}^{L} 2^{2r}\mathsf{Var}\left[\bar{\vartheta}_i z_{ir}\right] \qquad (122)
\end{aligned}$$

The last step follows since $z_{ir} \cdot z_{ir'} = 0$ whenever $r \neq r'$, since i may lie in only one sampled group.

Simplifying (122), we have,

$$\mathsf{Var}\left[\bar{\vartheta}_i z_{ir}\right] \leq \mathbb{E}\left[\bar{\vartheta}_i^2 z_{ir}\right] = \mathbb{E}\left[\bar{\vartheta}_i^2 \mid z_{ir} = 1\right]\mathsf{Pr}\left[z_{ir} = 1\right] \qquad (123)$$

Assuming that r is a level that is consistent with i (otherwise $\mathsf{Pr}\left[z_{ir} = 1 \mid \mathcal{G}\right] = 0$), we have, by Lemma 13 that $\mathbb{E}\left[\bar{\vartheta}_i \mid \mathcal{G}\right] \in |f_i|^p(1 \pm \delta)$ where, $\eta_{i,j,l_d(i)} \leq |f_i|/(15p)$, for $j \in R_l(i)$. Using Lemma 6, we obtain,

$$\begin{aligned}
\mathbb{E}\left[\bar{\vartheta}_i^2 \mid z_{ir} = 1, \mathcal{G}\right] &= \mathsf{Var}\left[\bar{\vartheta}_i \mid z_{ir} = 1, \mathcal{G}\right] + \left(\mathbb{E}\left[\bar{\vartheta}_i \mid z_{ir} = 1, \mathcal{G}\right]\right)^2 \\
&\leq \left(\frac{(0.288)p^2}{k}\right)(|f_i|^{2p-2})\left(\frac{|f_i|^{2p}}{(15p)^2}\right) + |f_i|^{2p}(1+\delta) \\
&\leq \frac{|f_i|^{2p}}{(750)k} + |f_i|^{2p}(1+\delta) \\
&\leq |f_i|^{2p}(1.001) \qquad (124)
\end{aligned}$$

where, $\delta \leq n^{-2500p}$.

Substituting (124) and (123) into (122), we have,

$$\begin{aligned}
\mathsf{Var}\left[Y_i \mid \mathcal{G}\right] &\leq \sum_{r=0}^{L} 2^{2r}\mathbb{E}\left[\bar{\vartheta}_i^2 z_{ir} \mid \mathcal{G}\right] \\
&\leq |f_i|^{2p}(1.001)\sum_{r=0}^{L} 2^{2r}\mathsf{Pr}\left[i \in \bar{G}_r \mid \mathcal{G}\right] \\
&\leq 2^{l+1}(1.001)|f_i|^{2p}\sum_{r=0}^{L} 2^{r}\mathsf{Pr}\left[i \in \bar{G}_r \mid \mathcal{G}\right] \\
&\leq (1.001)2^{l+1}|f_i|^{2p}(1+\delta) \\
&\leq (1.002)2^{l+1}|f_i|^{2p} \qquad (125)
\end{aligned}$$

Step 2 uses (124). Step 3 uses Lemma 8 to argue that if $i \in G_l$, then, $\mathsf{Pr}\left[i \in \bar{G}_r \mid \mathcal{G}\right] = 0$ for all $r > l+1$. Hence, the summation from $r = 0$ to L is equivalent to r ranging over $l-1, l$ and $l+1$. So the term $2^{2r} \leq 2^{l+1}2^r$. The last step again uses Lemma 8 to note that $\sum_{r=0}^{L} 2^r \mathsf{Pr}\left[i \in \bar{G}_r \mid \mathcal{G}\right] = 1 \pm O(2^l n^{-c})$. □

G.3 Covariance of Y_i and Y_j

Proof of Lemma 18. Let $i \neq j$, $i \in G_l$ and $j \in G_m$.

$$\begin{aligned}
&\mathsf{Cov}\left(Y_i, Y_j \mid \mathcal{G}\right) = \mathbb{E}\left[Y_i Y_j \mid \mathcal{G}\right] - \mathbb{E}\left[Y_i \mid \mathcal{G}\right]\mathbb{E}\left[Y_j \mid \mathcal{G}\right] \\
&= \mathbb{E}\left[\sum_{r=0}^{L} 2^r z_{ir}\bar{\vartheta}_i \sum_{r'=0}^{L} 2^{r'} z_{jr'}\bar{\vartheta}_j \mid \mathcal{G}\right] - \mathbb{E}\left[\sum_{r=0}^{L} 2^r z_{ir}\bar{\vartheta}_i \mid \mathcal{G}\right]\mathbb{E}\left[\sum_{r'=0}^{L} 2^{r'} z_{jr'}\bar{\vartheta}_j \mid \mathcal{G}\right] \\
&= \sum_{0\leq r,r'\leq L} 2^{r+r'}\mathbb{E}\left[\bar{\vartheta}_i\bar{\vartheta}_j \mid z_{ir}=1, z_{jr'}=1, \mathcal{G}\right]\mathsf{Pr}\left[z_{ir}=1, z_{jr'}=1 \mid \mathcal{G}\right] \\
&\quad - 2^{r+r'}\mathbb{E}\left[\bar{\vartheta}_i \mid z_{ir}=1 \mid \mathcal{G}\right]\mathbb{E}\left[\bar{\vartheta}_j \mid z_{jr'}=1, \mathcal{G}\right]\mathsf{Pr}\left[z_{ir}=1 \mid \mathcal{G}\right]\mathsf{Pr}\left[z_{jr'}=1 \mid \mathcal{G}\right] \\
&= \sum_{0\leq r,r'\leq L} 2^{r+r'}\sum_{\hat{f}_i,\hat{f}_j}\mathbb{E}\Big[\bar{\vartheta}_i\bar{\vartheta}_j \mid \hat{f}_i, \hat{f}_j, z_{ir}=1, z_{jr'}=1, \mathcal{G}\Big] \\
&\qquad\qquad \cdot \mathsf{Pr}\left[\hat{f}_i, \hat{f}_j \mid z_{ir}=1, z_{jr'=1}, \mathcal{G}\right]\cdot\mathsf{Pr}\left[z_{ir}=1, z_{jr'}=1 \mid \mathcal{G}\right] \\
&\quad - 2^{r+r'}\left(\sum_{\hat{f}_i}\mathbb{E}\left[\bar{\vartheta}_i \mid \hat{f}_i, z_{ir}=1, \mathcal{G}\right]\mathsf{Pr}\left[\hat{f}_i \mid z_{ir}=1, \mathcal{G}\right]\mathsf{Pr}\left[z_{ir}=1 \mid \mathcal{G}\right]\right) \\
&\qquad \left(\sum_{\hat{f}_j}\mathbb{E}\left[\bar{\vartheta}_j \mid \hat{f}_j, z_{jr'}=1, \mathcal{G}\right]\mathsf{Pr}\left[\hat{f}_j \mid z_{jr'=1}, \mathcal{G}\right]\mathsf{Pr}\left[z_{jr'}=1 \mid \mathcal{G}\right]\right)\Bigg] \quad . \qquad (126)
\end{aligned}$$

By Lemma 15,

$$\begin{aligned}
&\mathbb{E}\left[\bar{\vartheta}_i\bar{\vartheta}_j \mid \hat{f}_i, \hat{f}_j, z_{ir}=1, z_{jr'}=1, \mathcal{G}\right] \\
&= \mathbb{E}\left[\bar{\vartheta}_i \mid \hat{f}_i, \hat{f}_j, z_{ir}=1, z_{jr'}=1, \mathcal{G}\right]\cdot\mathbb{E}\left[\bar{\vartheta}_j \mid \hat{f}_i, \hat{f}_j, z_{ir}=1, z_{jr'}=1, \mathcal{G}\right] \quad . \qquad (127)
\end{aligned}$$

By Lemma 14, for any value of $\hat{f}_i$ satisfying $\mathcal{G}$ and $i \in \bar{G}_r$ such that r is consistent with $|f_i|$, we have,

$$\mathbb{E}\big[\bar{\vartheta}_i \mid \hat{f}_i, z_{ir} = 1, E', \mathcal{G}\big] = |f_i|^p(1 \pm \delta)$$

where, E' is any subset (including the empty subset) of the events $\{\hat{f}_j \wedge z_{jr'} = 1\}$ and $\delta = O(n^{-2500p})$.

Substituting in (127) and for r, r' consistent with $|f_i|$ and $|f_j|$ respectively, we have,

$$\mathbb{E}\left[\bar{\vartheta}_i\bar{\vartheta}_j \mid \hat{f}_i, \hat{f}_j, z_{ir} = 1, z_{jr'} = 1, \mathcal{G}\right] = |f_i|^p|f_j|^p(1 \pm O(\delta))$$

In a similar manner, it follows that

$$\mathbb{E}\left[\bar{\vartheta}_i \mid \hat{f}_i, z_{ir} = 1, \mathcal{G}\right]\mathbb{E}\left[\bar{\vartheta}_j \mid \hat{f}_j, z_{jr} = 1, \mathcal{G}\right] = |f_i|^p|f_j|^p(1 \pm O(\delta))$$

Substituting these into (126), we have,

$$\begin{aligned}
&\mathbb{E}\left[Y_iY_j \mid \mathcal{G}\right] - \mathbb{E}\left[Y_i \mid \mathcal{G}\right]\mathbb{E}\left[Y_j \mid \mathcal{G}\right] \\
&= \sum_{\substack{0\le r,r'\le L \\ r,\,r' \text{ consistent} \\ \text{with } i,j \text{ resp.}}} \Bigg[2^{r+r'}|f_i|^p|f_j|^p(1 \pm O(\delta)) \sum_{\hat{f}_i,\hat{f}_j} \mathsf{Pr}\left[\hat{f}_i, \hat{f}_j \mid z_{ir} = 1, z_{jr'} = 1, \mathcal{G}\right] \cdot \mathsf{Pr}\left[z_{ir} = 1, z_{jr'} = 1 \mid \mathcal{G}\right] \\
&\qquad - 2^{r+r'}\Big(|f_i|^p|f_j|^p(1 \pm O(\delta))\Big)\Big(\sum_{\hat{f}_i} \mathsf{Pr}\left[\hat{f}_i \mid z_{ir} = 1\right] \mathsf{Pr}\left[z_{ir} = 1, \mathcal{G}\right]\Big) \\
&\qquad\qquad \cdot \Big(\sum_{\hat{f}_j} \mathsf{Pr}\left[\hat{f}_j \mid z_{jr'=1}, \mathcal{G}\right] \mathsf{Pr}\left[z_{jr'} = 1 \mid \mathcal{G}\right]\Big)\Bigg] \\
&= \sum_{\substack{0\le r,r'\le L \\ r,\,r' \text{ consistent} \\ \text{with } i,j \text{ resp.}}} \Big[2^{r+r'}|f_i|^p|f_j|^p(1 \pm O(\delta))\mathsf{Pr}\left[z_{ir} = 1, z_{jr'} = 1 \mid \mathcal{G}\right] \\
&\qquad - 2^{r+r'}|f_i|^p|f_j|^p(1 \pm O(\delta))\mathsf{Pr}\left[z_{ir} = 1 \mid \mathcal{G}\right]\mathsf{Pr}\left[z_{jr'} = 1 \mid \mathcal{G}\right]\Big]
\end{aligned} \tag{128}$$

since, each of the summations, namely, (a) $\sum_{\hat{f}_i,\hat{f}_j} \mathsf{Pr}\left[\hat{f}_i, \hat{f}_j \mid z_{ir} = 1, z_{jr'} = 1, \mathcal{G}\right]$, (b) $\sum_{\hat{f}_i} \mathsf{Pr}\big[\hat{f}_i \mid z_{ir} = 1, \mathcal{G}\big]$ and (c) $\sum_{\hat{f}_j} \mathsf{Pr}\left[\hat{f}_j \mid z_{jr'=1}, \mathcal{G}\right]$ are 1 respectively.

Further,

$$\begin{aligned}
\sum_{\substack{0\le r,r'\le L \\ r,\,r' \text{ consistent} \\ \text{with } i,j \text{ resp.}}} 2^{r+r'}\mathsf{Pr}\left[z_{ir} = 1, z_{jr'} = 1 \mid \mathcal{G}\right] &= \sum_{0\le r,r'\le L} 2^{r+r'}\mathsf{Pr}\left[z_{ir} = 1, z_{jr'} = 1 \mid \mathcal{G}\right] \\
&= 1 \pm O(2^l + 2^m)n^{-c}, \text{ by Lemma 46}
\end{aligned}$$

since, if levels r and r' are not consistent respectively with $|f_i|$ and $|f_j|$ respectively then $\mathsf{Pr}\big[z_{ir} = 1, z_{jr'} = 1 \mid \mathcal{G}\big] = 0$. The same applies to summations over r of $2^r\mathsf{Pr}\left[z_{ir} = 1 \mid \mathcal{G}\right]$, etc..

By Lemma 8 (part 4),

$$\sum_{r=0}^{L} 2^r\mathsf{Pr}\left[z_{ir} = 1 \mid \mathcal{G}\right] = \sum_{\substack{r \text{ consistent} \\ \text{with } i}} 2^r \cdot \mathsf{Pr}\left[z_{ir} = 1 \mid \mathcal{G}\right] = 1 \pm 2^l n^{-c} \ .$$

Similarly, $\sum_{r' \text{ consistent with } j} 2^{r'} \cdot \mathsf{Pr}\left[j \in \bar{G}_{r'}\right] \in 1 \pm 2^m n^{-c}$. Combining and taking absolute values of both sides in (128) and replacing equality by $\leq$, we have,

$$\begin{aligned}
&\left|\mathsf{Cov}\left(Y_i, Y_j \mid \mathcal{G}\right)\right| \\
&\leq |f_i|^p |f_j|^p \left((1 \pm O(\delta))(1 \pm O(2^l + 2^m)n^{-c}) - (1 \pm O(\delta))(1 \pm O(2^l n^{-c}))(1 \pm O(2^m n^{-c}))\right) \\
&= |f_i|^p |f_j|^p O(\delta + (2^l + 2^m)n^{-c}) \\
&= |f_i|^p |f_j|^p \cdot O(n^{-c+1})
\end{aligned}$$

□

G.4 Variance of $\hat{F}_p$ estimator

Proof of Lemma 19. Let $K = 425$, so that $B = Kn^{1-2/p}\epsilon^{-2}/\min(\log(n), \epsilon^{4/p-2}) \geq Kn^{1-2/p}\epsilon^{-4/p}$. We have,

$$\begin{aligned}
\mathsf{Var}\left[\hat{F}_p\right] &= \mathsf{Var}\left[\sum_{i\in[n]} Y_i \mid \mathcal{G}\right] \\
&\leq \sum_{i\in[n]} \mathsf{Var}\left[Y_i \mid \mathcal{G}\right] + \sum_{i\neq j} \left|\mathsf{Cov}\left(Y_i, Y_j \mid \mathcal{G}\right)\right| \\
&\leq \sum_{i\in \mathrm{mid}(G_0)} \mathsf{Var}\left[Y_i \mid \mathcal{G}\right] + \sum_{i\in[n], i\notin \mathrm{mid}(G_0)} \mathsf{Var}\left[Y_i \mid \mathcal{G}\right] + F_p^2 \cdot O(n^{-c+1}) \\
&\leq \sum_{i\in \mathrm{mid}(G_0)} \frac{(0.3)\epsilon^2 |f_i|^{2p-2} F_p^{2/p}}{(10)^4 K} + \left(\sum_{l=0}^{L} \sum_{i\in G_l, i\notin \mathrm{mid}(G_0)} 2^{l+1}(1.002)|f_i|^{2p}\right) + O(n^{-c+2})F_p^2
\end{aligned} \tag{129}$$

Step 3 follows from Lemma 18, since,

$$\sum_{i\neq j} \left|\mathsf{Cov}\left(Y_i, Y_j \mid \mathcal{G}\right)\right| \leq \sum_{i\neq j} O(n^{-c+1})|f_i|^p |f_j|^p \leq O(n^{-c+1}F_p^2) \ .$$

Step 4 uses Lemma 17. Since, $\hat{F}_2 \leq F_2(1 + 0.01/(2p))$, $\hat{F}_2^{p/2} \leq (1.01)F_2$. Also, $F_2 \leq F_p^{2/p} n^{1-2/p}$. Therefore,

$$\left(\hat{F}_2/B\right)^{p/2} \leq \left(\frac{(1.01)F_2}{Kn^{1-2/p}\epsilon^{-4/p}}\right)^{p/2} \leq (1.01/K)^{p/2}\epsilon^2 F_p \ . \tag{130}$$

For any set $S \subset [n]$ and $q \geq 0$, let $F_q(S)$ denote $\sum_{i\in S}|f_i|^q$.
Let $i \in \mathrm{lmargin}(G_0) \cup_{l=1}^{L} G_l$. By definitions of the parameters,

$$|f_i| \leq T_{l-1} \leq \left(\frac{F_2\left(1 + \frac{0.01}{2p}\right)}{(2\alpha)^{l-1}B}\right)^{1/2} \qquad \text{if } i \in G_l \text{ and } l \geq 1,$$

$$|f_i| \leq T_0(1+\bar{\epsilon}) \leq \left(\frac{F_2\left(1 + \frac{0.01}{2p}\right)}{B}\right)^{1/2} \left(1 + \frac{1}{27p}\right) \qquad \text{if } i \in \mathrm{lmargin}(G_0) \ .$$

We consider the first summation term of Eqn. (129), that is,

$$\sum_{i\in \text{mid}(G_0)} \frac{(0.3)\epsilon^2 |f_i|^{2p-2} F_p^{2/p}}{(10)^4 K} = \left(\frac{(0.3)\epsilon^2 F_p^{2/p}}{((10)^4)(425)}\right) F_{2p-2}(\text{mid}(G_0))$$
$$\leq \left(\frac{(0.3)\epsilon^2 F_p^{2/p}}{((10)^4)(425)}\right) F_p^{2-2/p}(\text{mid}(G_0)) \leq \frac{(0.3)\epsilon^2 F_p^2}{((10)^4)(425)} \quad (131)$$

We now consider the second summation term of Eqn. (129), that is,

$$\sum_{l=0}^{L} \sum_{i\in G_l, i\notin \text{mid}(G_0)} 2^{l+1}(1.002)|f_i|^{2p}$$
$$\leq \sum_{i\in \text{lmargin}(G_0)} (2)(1.002)(T_0(1+\bar{\epsilon}))^p |f_i|^p + \sum_{l=1}^{L}\sum_{i\in G_l} 2^{l+1} T_{l-1}^p |f_i|^p \quad (132)$$

We will consider the two summations in Eqn. (132) separately.

$$\sum_{i\in \text{lmargin}(G_0)} (2)(1.002)(T_0(1+\bar{\epsilon}))^p |f_i|^p$$
$$\leq (2)(1.002)\left(\frac{F_2}{B}\right)^{p/2} (1.01)e^{1/27} F_p(\text{lmargin}(G_0))$$
$$\leq (2.11)\left(\frac{n^{1-2/p}F_p^{2/p}}{(425)n^{1-2/p}\epsilon^{-4/p}}\right)^{p/2} F_p(\text{lmargin}(G_0))$$
$$= \left(\frac{1}{200}\right)\epsilon^2 F_p \cdot F_p(\text{lmargin}(G_0)) \quad (133)$$

We now consider the second summation of Eqn. (132).

$$\sum_{l=1}^{L}\sum_{i\in G_l} 2^{l+1} T_{l-1}^p |f_i|^p$$
$$\leq (2)(1.01)\sum_{l=1}^{L} 2^l \left(\frac{F_2}{(2\alpha)^{l-1}B}\right)^{p/2} F_p(G_l)$$
$$= (4)(1.01)\sum_{l=1}^{L} 2^l \left(\frac{F_p^{2/p} n^{1-2/p}}{(425)(2\alpha)^{l-1}n^{1-2/p}\epsilon^{-4/p}}\right)^{p/2} F_p(G_l)$$
$$= \left(\frac{(4.04)}{(425)^{p/2}}\right)\epsilon^2 F_p \sum_{l=1}^{L} 2^l (2\alpha)^{-(l-1)(p/2)} F_p(G_l) \quad (134)$$

Further,

$$2^l(2\alpha)^{(l-1)(-p/2)} = (2\alpha)^{p/2} 2^l (2\alpha)^{-lp/2} = (2\alpha)^{p/2} 2^{l(1-(p/2)\log_2(2\alpha))} \ . \quad (135)$$

Let $\gamma = 1 - \alpha = (1 - 2/p)\nu$, where, $\nu = 0.01$. Therefore,

$$\begin{aligned}\log_2(2\alpha) &= 1 + \log_2(\alpha) = 1 + \frac{\ln(\alpha)}{\ln 2} = 1 + \frac{\ln(1-\gamma)}{\ln 2} \\ &\geq 1 - \frac{2\gamma}{\ln 2} = 1 - \frac{2(1-2/p)\nu}{\ln 2} \geq 1 - (1-2/p)(3\nu) \end{aligned} \tag{136}$$

Using eqn. (136), we can simplify the term $1 - (p/2)\log_2(2\alpha)$ as

$$1 - (p/2)\log_2(2\alpha) \leq 1 - (p/2)\left(1 - (1-2/p)(3\nu)\right) = -(p/2-1)(1-3\nu) = -(p/2-1)(0.97) < 0$$

and is a constant.

Substituting this into Eqn. (135) and then into (134), we have,

$$\begin{aligned}&\sum_{l=1}^{L}\sum_{i\in G_l} 2^{l+1}T_{l-1}^{p}|f_i|^p \\ &\leq \left(\frac{(4.04)}{(425)^{p/2}}\right)\epsilon^2 F_p \sum_{l=1}^{L} 2^l (2\alpha)^{-(l-1)(p/2)} F_p(G_l) \\ &\leq (4.04)\left(\frac{2\alpha}{425}\right)^{p/2}\epsilon^2 F_p \sum_{l=1}^{L}\left(2^{-(p/2-1)(1-3\nu)}\right)^l F_p(G_l) \\ &\leq \left(\frac{\epsilon^2}{53}\right) F_p \sum_{l=1}^{L} F_p(G_l) \\ &\leq \left(\frac{\epsilon^2}{53}\right) F_p \cdot F_p(\cup_{l=1}^{L} G_l)\ . \end{aligned} \tag{137}$$

Adding Eqns. (133) and (137), Eqn. (132) becomes

$$\begin{aligned}&\sum_{l=0}^{L}\sum_{i\in G_l, i\notin \mathrm{mid}(G_0)} 2^{l+1}(1.002)|f_i|^{2p} \\ &\leq \left(\frac{1}{200}\right)\epsilon^2 F_p \cdot F_p(\mathrm{lmargin}(G_0)) + \left(\frac{\epsilon^2}{53}\right) F_p \cdot F_p(\cup_{l=1}^{L} G_l) \\ &\leq \frac{\epsilon^2 F_p^2}{53} \end{aligned} \tag{138}$$

Substituting Eqn. (131) and Eqn. (138) in Eqn. (129), we have, for n sufficiently large, that

$$\mathsf{Var}\left[\hat{F}_p\right] \leq \frac{\epsilon^2 F_p^2}{50}$$

□

G.5 Putting things together

Proof of Theorem 20. Consider the Geometric-Hss algorithm using the parameters of Figure 2. By Lemma 7, $\mathcal{G}$ holds except with probability n^{-c}, where, $c > 23$. From Lemma 19, $\mathsf{Var}\left[\hat{F}_p\right] \leq$

$\epsilon^2 F_p^2/50$. Using Chebychev's inequality,

$$\mathsf{Pr}\left[\left|\hat{F}_p - \mathbb{E}\left[\hat{F}_p\right]\right| \leq (\epsilon/2)F_p \mid \mathcal{G}\right] \geq 1 - \frac{\mathsf{Var}\left[F_p\right]}{((\epsilon/2)F_p)^2} = 1 - \frac{4}{50} \ . \tag{139}$$

By Lemma 16, $\left|\mathbb{E}\left[\hat{F}_p \mid \mathcal{G}\right] - F_p\right| \leq F_p(2^{L+1}n^{-c})$. Combining with Eqn. (139), by triangle inequality, we have,

$$\mathsf{Pr}\left[\left|\hat{F}_p - F_p\right| \leq \left((\epsilon/2) + 2^{L+1}n^{-c})\right) F_p \mid \mathcal{G}\right] \geq 1 - \frac{4}{50}$$

which implies that

$$\mathsf{Pr}\left[\left|\hat{F}_p - F_p\right| \leq \epsilon F_p \mid \mathcal{G}\right] \leq \frac{46}{50} \ .$$

since, $2^L \ll n$.
Since $\mathsf{Pr}\left[\mathcal{G}\right] \geq 1 - n^{-c}$, unconditioning w.r.t. $\mathcal{G}$, we have,

$$\mathsf{Pr}\left[\left|\hat{F}_p - F_p\right| \leq \epsilon F_p\right] \geq \frac{46}{50}\left(1 - O(n^{-c})\right) \geq 0.9 \ .$$

The space required at level 0 is $C_0 s = Cs$, at level l it is $C_l s$ and at level L it is $16C_L s$. Here, $s = 8k = 8(1000)\log(n) = O(\log n)$. Further $C_l = 4(\alpha)^l C$. Thus, the total space is of the order of

$$\sum_{l=0}^{L} C_l s = (\log(n)) \sum_{l=0}^{L} \alpha^l C \leq \frac{C\log(n)}{1-\alpha} = \frac{C\log(n)}{(1-2/p)\nu} = O\left(\frac{n^{1-2/p}\log(n)\epsilon^{-2}}{\min(\log(n), \epsilon^{4/p-2})}\right) \ .$$

The last expression for space may also be written as $O\left(n^{1-2/p}\epsilon^{-2} + n^{1-2/p}\epsilon^{-4/p}\log(n)\right)$.

The time taken to process each stream update consists of applying the L hash functions $g_1, \ldots, g_L$ to an item i. Each hash function is $O(\log n)$-wise independent and requires time $O(\log n)$ to evaluate it at a point. The time to evaluate $L = \log_{2\alpha}(n/C)$ functions is $O(\log^2 n)$. Additionally, for each level, the hash values for i have to be computed for each of the s hash functions of the HH_l and $\textsc{tpest}_l$ structures. These hash functions are $O(1)$-wise independent, and they can collectively be computed in $O(Ls) = O(\log^2 n)$ time. This proves the statement of the theorem.

□